LiveVBR: Energy-efficient Live Per-title Encoding for Adaptive Video Streaming

Vignesh V Menon[1], **Prajit T Rajendran**[2]
[1]Christian Doppler Laboratory ATHENA, Alpen-Adria-Universität, Klagenfurt, Austria
[2]Universite Paris-Saclay, CEA, List, F-91120, Palaiseau, France

30 June 2023

Outline

1. Research Goal
2. Live-VBR
3. Results
4. Summary

Research Goal

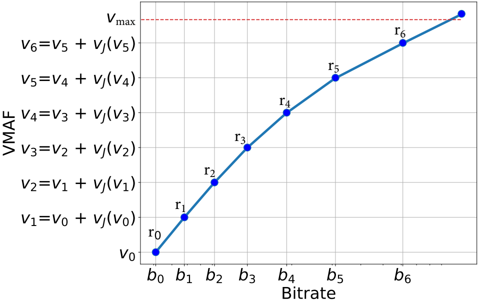


Figure: The ideal perceptually-aware bitrate ladder envisioned in this work. Here, $v_J(v_0) = v_J(v_1) = v_J(v_{M-1}) = \Delta$VMAF

Joint optimization:

- Perceptual difference of pre-defined v_J between representations.
- Minimize bitrate difference between representations.
- Maximize compression efficiency of representations.

Workflow of Live-VBR

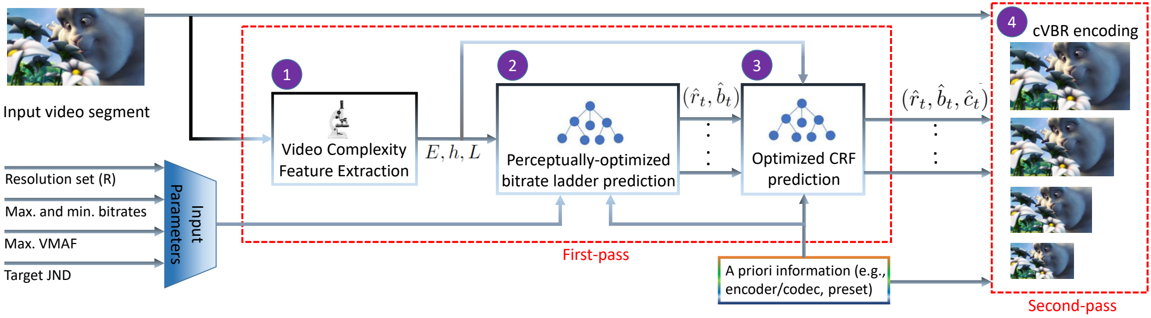
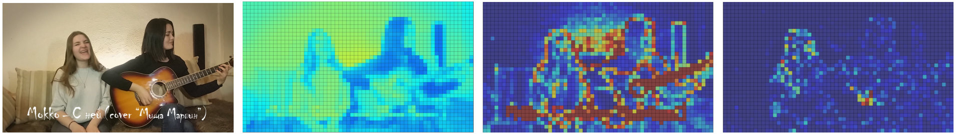


Figure: Live HTTP adaptive streaming featuring Live-VBR.

Video Complexity Feature Extraction

- Accomplished using VCA.[1]
  - E: the average texture energy
  - h: the average gradient of the texture energy
  - L: the average luminescence



(a) Original frame (b) Heatmap of L (c) Heatmap of E (d) Heatmap of h

Figure: Example heatmap of Luminescence (L), spatial texture (E) and temporal activity (h) features of the 2nd frame of *CoverSong_1080P_0a86* video of Youtube UGC dataset extracted using VCA.

[1] Vignesh V Menon et al. "VCA: Video Complexity Analyzer". In: *Proceedings of the 13th ACM Multimedia Systems Conference*. MMSys '22. Athlone, Ireland: Association for Computing Machinery, 2022, 259–264. ISBN: 9781450392839. DOI: 10.1145/3524273.3532896. URL: https://doi.org/10.1145/3524273.3532896.

Live-VBR

First point of the bitrate ladder[2]

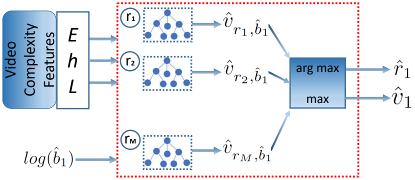


Figure: Estimation of the first point of the bitrate ladder.

$\hat{b}_1 = b_{min}$
Determine $\hat{v}_{r,\hat{b}_1}$ $\forall r \in R$
$\hat{v}_1 = max(\hat{v}_{r,\hat{b}_1})$
$\hat{r}_1 = \arg\max_{r\in R}(\hat{v}_{r,\hat{b}_1})$
$(\hat{r}_1, \hat{b}_1)$ is the first point of the bitrate ladder.

[2] V. V. Menon et al. "OPTE: Online Per-Title Encoding for Live Video Streaming". In: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2022, pp. 1865–1869. DOI: 10.1109/ICASSP43922.2022.9746745.

Live-VBR

Remaining points of the bitrate ladder[3]

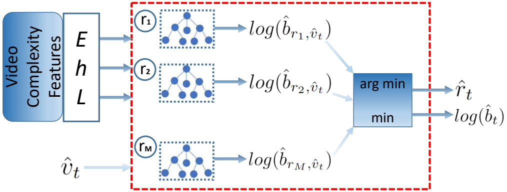


Figure: Estimation of the t^{th} point of the bitrate ladder.

$t = 2$

while $\hat{b}_{t-1} < b_{max}$ *and* $\hat{v}_{t-1} < v_{max}$ **do**

- $\hat{v}_t = \hat{v}_{t-1} + v_J(\hat{v}_{t-1})$
- Determine $\hat{b}_{r,\hat{v}_t}$ $\forall r \in R$
- $\hat{b}_t = min(\hat{b}_{r,\hat{v}_t})$
- $\hat{r}_t = \arg\min_{r \in R}(\hat{b}_{r,\hat{v}_t})$
- $(\hat{r}_t, \hat{b}_t)$ is the t^{th} point of the bitrate ladder.
- $t = t + 1$

[3] V. V. Menon et al. "Perceptually-Aware Per-Title Encoding for Adaptive Video Streaming". In: *2022 IEEE International Conference on Multimedia and Expo (ICME)*. Los Alamitos, CA, USA: IEEE Computer Society, 2022, pp. 1–6. DOI: 10.1109/ICME52920.2022.9859744. URL: https://doi.ieeecomputersociety.org/10.1109/ICME52920.2022.9859744.

Live-VBR

cVBR encoding of the bitrate ladder[4]

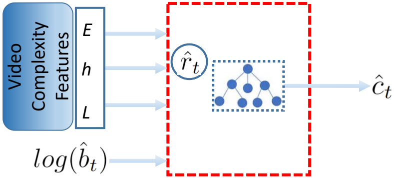


Figure: Estimation of the optimized CRF to achieve the target bitrate $\hat{b}_t$ using a prediction model trained for resolution $\hat{r}_t$.

- Optimized CRF is determined for the selected $(\hat{r}_t, \hat{b}_t)$ pairs.
- cVBR encoding for the $(\hat{r}_t, \hat{b}_t, \hat{c}_t)$ pairs is performed.

[4] Vignesh V Menon et al. "ETPS: Efficient Two-Pass Encoding Scheme for Adaptive Live Streaming". In: *2022 IEEE International Conference on Image Processing (ICIP)*. 2022, pp. 1516–1520. DOI: 10.1109/ICIP46576.2022.9897768.

Prediction models

Design

- **VMAF Prediction models**
  - Input: Video complexity features (E, h, L), target bitrate $\hat{b}_t$
  - Output: VMAF score
- **Bitrate Prediction models**
  - Input: Video complexity features (E, h, L), predicted VMAF $\hat{v}_t$
  - Output: Predicted bitrate $\hat{b}_t$
- **CRF Prediction models**
  - Input: Video complexity features (E, h, L), target bitrate $\hat{b}_t$
  - Output: CRF value

Prediction models

Architecture and training

- Implementation using Python
- Dataset used for training: VCD dataset[5] @ 30 fps
- Train-validation-test split: 70-10-20, clustered based on feature space
- Prediction model architecture: Random forest (RF) models
- Advantages of RF models: Good prediction performance, lower variance, robust to outlier
- Modular approach to ensure scalability

[5] Hadi Amirpour et al. "VCD: Video Complexity Dataset". In: *Proceedings of the 13th ACM Multimedia Systems Conference*. MMSys '22. Athlone, Ireland: Association for Computing Machinery, 2022, 234–239. ISBN: 9781450392839. DOI: 10.1145/3524273.3532892. URL: https://doi.org/10.1145/3524273.3532892.

Results

Prediction accuracy of the models

Table: R^2 score and MAE of the prediction models for various resolutions.

	R^2 score							MAE						
r	360p	432p	540p	720p	1080p	1440p	2160p	360p	432p	540p	720p	1080p	1440p	2160p
VMAF	0.821	0.852	0.882	0.906	0.910	0.906	0.930	5.091	5.071	4.966	4.971	4.806	4.490	3.941
log(b)	0.867	0.884	0.901	0.910	0.932	0.937	0.943	0.527	0.505	0.472	0.456	0.460	0.472	0.488
CRF	0.969	0.969	0.970	0.969	0.968	0.967	0.968	1.823	1.820	1.820	1.859	1.860	1.885	1.871

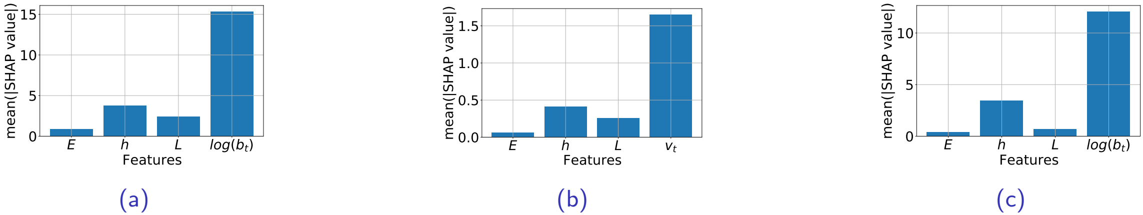


Figure: The relative importance of (a) VMAF prediction, (b) bitrate prediction, and (c) CRF prediction for 2160p resolution determined by SHAP values.

Results

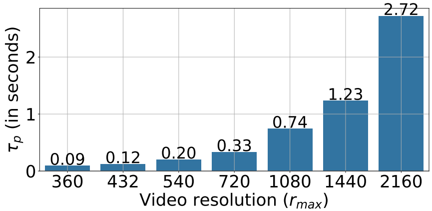
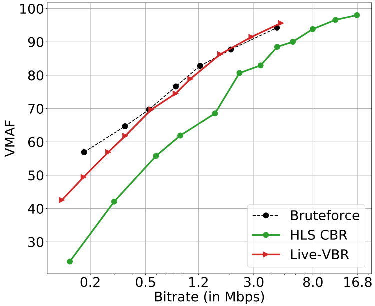


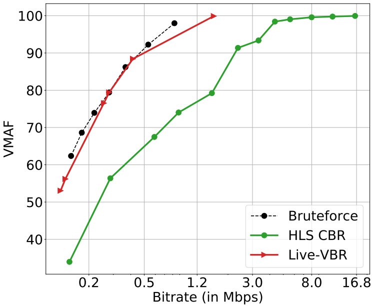
Figure: Pre-processing time (τ_p) of Live-VBR for various input video resolutions.

Results

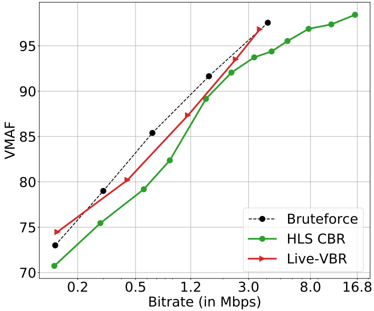
RD plots of Live-VBR using x265



(a) *Bunny_s000*



(b) *Characters_s000*

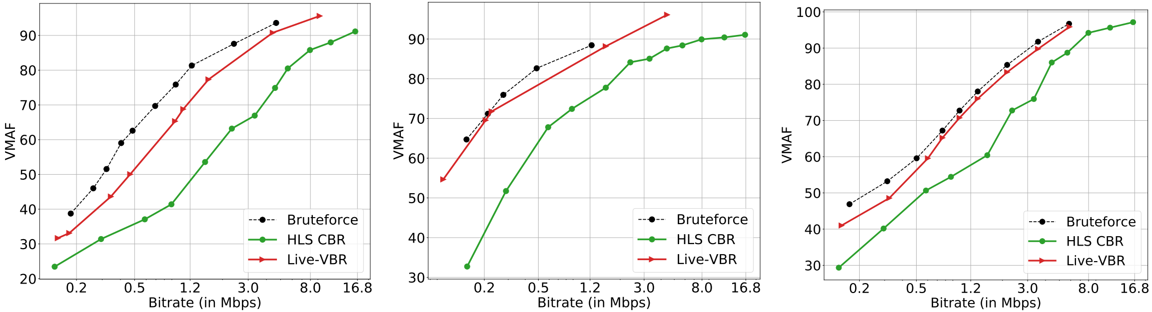


(c) *Eldorado_s000*

Figure: RD curves of representative video sequences (segments) (a) *Bunny_s000* (E =22.40, h=4.70, L=129.21), (b) *Characters_s000* (E =45.42, h=36.88, L=134.56), (c) *Eldorado_s000* (E=15.28, h=49.76, L=140.54) using the HLS CBR encoding (green line), and Live-VBR encoding (red line). JND is considered as 6 VMAF in these plots.

Results

RD plots of Live-VBR using x265



(a) *Eldorado_s005* (b) *HoneyBee_s000* (c) *Wood_s000*

Figure: RD curves of representative video sequences (segments) (a) *Eldorado_s005* (E =100.37, h=9.23, L=109.06), (b) *HoneyBee_s000* (E=42.93, h=7.91, L=103.00), (c) *Wood_s000* (E=124.72, h=47.03, L=119.57) using the HLS CBR encoding (green line), and Live-VBR encoding (red line). JND is considered as 6 VMAF in these plots.

Results

Table: Average results of the encoding schemes compared to the HLS CBR encoding.

Method	BDR_P	BDR_V	BD-PSNR	BD-VMAF	ΔS	ΔT
Bruteforce (2 VMAF JND)[6]	-23.09%	-43.23%	1.34 dB	10.61	-25.99%	4732.33%
Bruteforce (4 VMAF JND)	-28.15%	-42.75%	1.70 dB	10.08	-59.07%	4732.33%
Bruteforce (6 VMAF JND)	-25.36%	-40.73%	1.67 dB	9.19	-70.50%	4732.33%
Live-VBR (2 VMAF JND)	-14.25%	-29.14%	1.36 dB	7.82	23.57%	184.62%
Live-VBR (4 VMAF JND)	-18.41%	-32.48%	1.41 dB	8.31	-56.38%	26.14%
Live-VBR (6 VMAF JND)	-18.80%	-32.59%	1.34 dB	8.34	-68.96%	-18.58%

Relative storage difference

$$\Delta S = \frac{\sum b_{opt}}{\sum b_{ref}} - 1$$

Relative difference in the encoding time

$$\Delta T = \frac{\tau_p + \sum t_{opt}}{\sum t_{ref}} - 1$$

[6] Jan De Cock et al. "Complexity-based consistent-quality encoding in the cloud". In: *2016 IEEE International Conference on Image Processing (ICIP)*. 2016, pp. 1484–1488. DOI: 10.1109/ICIP.2016.7532605.

Summary and Future Directions

- Introduced an optimized encoding bitrate ladder prediction scheme, which uses RF-based models to estimate bitrate-resolution-CRF triples for a given video segment based on its spatial and temporal characteristics.
- The bitrate ladder is predicted such that there is a perceptual difference of at least one JND between the representations.
- The prototype implementation code is released at: `https://github.com/cd-athena/LiveVBR`.
- The online documentation can be found at: `https://cd-athena.github.io/LiveVBR/`.

Further details:

Vignesh V Menon, Prajit T Rajendran, Christian Feldmann, Klaus Schoeffmann, Mohammad Ghanbari and Christian Timmerer, "JND-aware Two-pass Per-title Encoding Scheme for Adaptive Live Streaming," in IEEE Transactions on Circuits and Systems for Video Technology, doi: 10.1109/TCSVT.2023.3290725.

Q & A

Thank you for your attention!

Vignesh V Menon (vignesh.menon@aau.at)
Prajit T Rajendran (prajit.thazhurazhikath@cea.fr)