

ALL-INTRA RATE CONTROL USING LOW COMPLEXITY VIDEO FEATURES FOR VERSATILE VIDEO CODING

Vignesh V Menon¹ Anastasia Henkel² Prajit T Rajendran³ Christian R. Helmrich²
Adam Wieckowski² Benjamin Bross² Christian Timmerer¹ Detlev Marpe²

¹ Christian Doppler Laboratory ATHENA, Alpen-Adria-Universität, Klagenfurt, Austria

² Video Communication and Applications Department, Fraunhofer HHI, Berlin, Germany

³ CEA, List, F-91120 Palaiseau, Université Paris-Saclay, France

ABSTRACT

Versatile Video Coding (VVC) allows for large compression efficiency gains over its predecessor, *High Efficiency Video Coding* (HEVC). The added efficiency comes at the cost of increased runtime complexity, especially for encoding. It is thus highly relevant to explore all available runtime reduction options. This paper proposes a novel first pass for two-pass rate control in all-intra configuration, using low-complexity video analysis and a Random Forest (RF)-based machine learning model to derive the data required for driving the second pass. The proposed method is validated using VVenC, an open and optimized VVC encoder. Compared to the default two-pass rate control algorithm in VVenC, the proposed method achieves around 32% reduction in encoding time for the preset *faster*, while on average only causing 2% BD-rate increase and achieving similar rate control accuracy.

Index Terms— Rate control, Complexity reduction, Random Forest, Machine learning, VVC.

1. INTRODUCTION

Modern video standards come with ever-increasing complexity. The newest, *Versatile Video Coding* (VVC) [1], was already during its planning intended to be up to ten times more complex to encode than its predecessor, *High Efficiency Video Coding* (HEVC) [2]. Practical implementations, like the open source Versatile Video Encoder (VVenC) [3], can efficiently deal with scaling the compression efficiency versus the runtime [4] as shown in Fig. 1, by providing different presets allowing to trade-off runtime against compression efficiency.

Motivation: With reduced runtime complexity, additional processing steps of constant runtime are increasingly significant. In VVenC, especially the rate control shows this behavior. The encoder processes the input signal twice for the two-pass rate control (2pRC) [5,6]. The first pass, *i.e.*, encoding the video using a fixed reduced toolset, is used to collect basic statistics. Those are then used in the second pass, using the desired working point (or preset), to drive bit allocation between pictures for optimal rate distribution (*cf.* Fig. 2a)

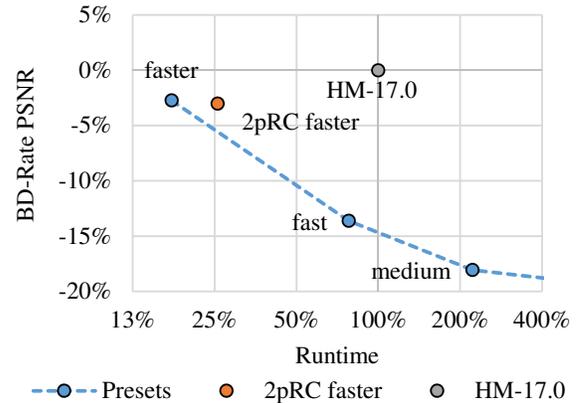


Fig. 1: VVenC 1.7.0 presets for fixed QP all-intra encoding as well as the working point of two-pass rate control for the preset *faster*, compared to HM-17.0.

such that (i) the overall quality of the video is constant over time, and (ii) the final encoding has roughly a specific size. While for the high-efficiency presets (*e.g.*, *slower* preset) of VVenC, the 2pRC encoding does not take substantially longer than an encoding using a fixed quantization parameter (QP) value resulting in a specific rate, the overhead of the first pass is higher for faster presets. However, for the preset *faster*, the 2pRC runtime is up to 150% of a fixed QP encoding resulting in a similar rate. In Fig. 1, it can be observed that the runtime of 2pRC for preset *faster* lies significantly below the Pareto front at a comparable speed.

All-Intra video coding is a video compression method that encodes each video frame independently, without referring to any previously encoded frames. This is in contrast to the more common inter-frame coding methods, where the encoding of a frame is based on the difference between the current frame and previously encoded frames. All-intra coding is vital because it provides high-quality video, low latency, improved error resilience, random access, and efficient editing. These benefits make all-intra coding useful in live video streaming and professional video production applications. In all-intra encoding, no motion compensation or other inter-frame de-

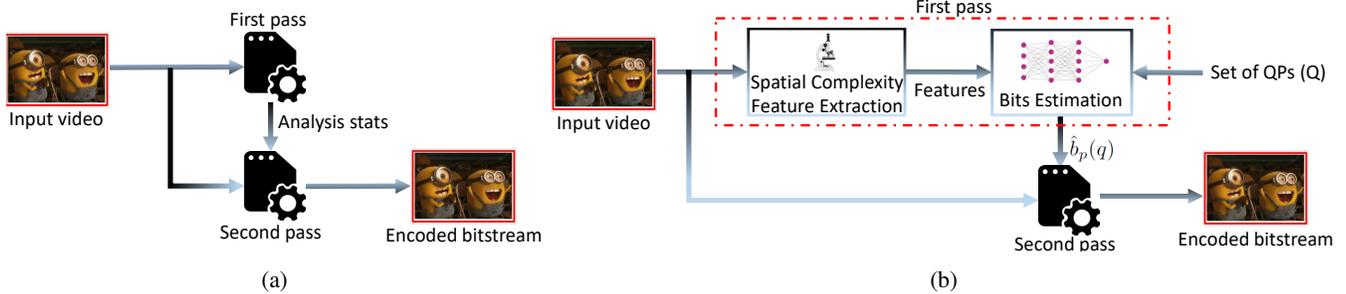


Fig. 2: (a) State-of-the-art and (b) proposed two-pass rate control encoding architecture.

dependencies are allowed. For example, such a mode can be used as a benchmark for motion-compensated modes [7]. All-intra coding is also widely used in practice, *e.g.*, when encoding single frames (*i.e.*, still picture coding) or when instant random access to every frame is required (*e.g.*, mezzanine codecs are mostly all-intra). In this paper, the experiments are performed for all-intra coding conditions.

Target: This paper aims to minimize the first pass encoding time, *i.e.*, time taken to collect statistics used in the second pass while achieving similar rate control accuracy. To this light, this paper explores an alternative method for statistics collection during the first pass to reduce the overall runtime of two-pass rate control encoding, one of the essential encoding modes in state-of-the-art video encoders like x264¹, x265², and VVenC [3]. As can be observed in [4] for the default use case of motion compensated encoding using VVenC, and in [8] for other use cases, a runtime increase of 50% around the preset *faster* provides a substantial bitrate reduction if following the Pareto front.

Contributions: This paper proposes a two-pass rate control method for intra-coding that includes using a light-weight estimator of frame complexity as the first pass as shown in Fig. 2b. The proposed first pass produces the statistics used in the second encoding pass with lower computational overhead than the default two-pass encoding scheme in the state-of-the-art video encoders. Video Coding Analyzer (VCA) [9] is used as the low-complexity estimator. While VCA has proven helpful in estimating encoding complexity [10–13], it does not provide the exact statistics as input to the second pass encoding rate control, *i.e.*, the bitrate distribution from the first pass. A Random Forest (RF)-based machine learning model is designed to predict the required number of bits from the VCA complexity estimation to overcome this problem. The proposed two-pass rate control method is validated using VVenC. It achieves around 32% reduction in encoding time for the preset *faster*, while on average only causing 2% BD-rate increase and achieving similar rate control accuracy compared to the two-pass rate control method in VVenC.

Paper outline: Section 2 describes the proposed operation of a reduced complexity two-pass rate control for VVenC. In Section 3, the accuracy of the designed model and the empirical results of the encoding system described in Section 2 are evaluated. Section 4 concludes the paper.

2. TWO-PASS RATE CONTROL USING LOW-COMPLEXITY BIT ESTIMATION

2.1. First Pass

The first pass of the proposed rate control method is divided into three steps: (i) spatial complexity feature extraction, (ii) bits estimation, and (iii) application to VVenC.

(i) *Spatial complexity feature extraction:* The commonly used spatial complexity feature is Spatial Information (SI) [14], but its correlation with encoding output features such as bitrate, encoding time, *etc.* is very low, which is insufficient for encoding parameter prediction [9]. In this paper, six DCT-energy-based features are used:

- the average luminance texture energy E_Y ;
- the average luminance L_Y ;
- the average chrominance texture energy E_U and E_V (for U and V planes); and
- the average chrominance L_U and L_V (for U and V planes).

These features are extracted using the open source Video Complexity Analyzer (VCA)³ [9] and represented as the following vector.

$$x = [E_Y, L_Y, E_U, L_U, E_V, L_V] \quad (1)$$

(ii) *Bits estimation:* For each I-frame, the number of bits is predicted using the spatial features (*i.e.*, luminance and chrominance features) of the frame for each quantization parameter q . In this paper, a random forest regression model is used.

$$\tilde{x} = [x|q]^T \quad (2)$$

The predicted bits \hat{b} can be presented as $\hat{b} = f(\tilde{x})$. The loss function used for training this model is the mean squared error

¹<https://www.videolan.org/developers/x264.html>, last access: Feb 20, 2023.

²<https://www.videolan.org/developers/x265.html>, last access: Feb 20, 2023.

³<https://vca.itec.aau.at>, last access: Feb 15, 2023.

(MSE), which measures the average difference between the predicted and actual (ground truth) number of bits.

(iii) *Application to VVenC*: For each frame p , the above bits estimator (ii) provides a bit-count prediction \hat{b}_p for each possible QP value q_p . Then, the first pass compiles a list of \hat{b}_p values for each p according to the q_p values assigned to each frame in the first pass, using $f(\tilde{x})$.

2.2. Second Pass

For each frame, the first pass rate-QP estimator of Section 2.1 provides a pre-assigned QP value q_p and an associated bit-count prediction \hat{b}_p . Using this data pair and the target bit count b'_p for the second pass, VVenC's *R-QP* model [5] determines the closest integer QP value q'_p corresponding to b'_p and performs the second pass encoding using that q'_p :

$$\bar{q}_p = q_p - c_{\text{low}} \cdot \sqrt{\max(1; q_p)} \cdot \log_2 \left(\frac{b'_p}{\hat{b}_p} \right), \quad (3)$$

where c_{low} is a constant for the low-rate end of the *R-QP* function and \bar{q}_p is an initial second pass QP value. The final q'_p is obtained using a high-rate corrective step as:

$$q'_p = \text{round}(\bar{q}_p + c_{\text{high}} \cdot \max(0; q_{\text{start}} - \bar{q}_p)), \quad (4)$$

where $0 < c_{\text{high}} < 1$ is a video resolution dependent constant (e.g., 0.5 for 2160p and 0.25 for 480p input) and $q_{\text{start}} = 24$ was chosen experimentally. Further details can be found in [15].

Any inaccuracies in the bit consumption of a frame after this final rate-distortion optimal encoding pass (i.e., too many or too few bits spent in the final pass) will be accumulated and compensated for in the following frames, as described in [5].

3. EVALUATION

3.1. Bits Estimation Model

Training and Hyperparameters: The hyperparameters used in the RF model are *random_state* = 0, *min_samples_leaf* = 1, *min_samples_split* = 2, and *n_estimators* = 100. For *max_depth*, four different values, i.e., 4, 8, 12, and 16, have been experimented with, considering a trade-off between model size and prediction accuracy. In this paper, prediction accuracy is evaluated using the coefficient of determination (R^2) score and Mean Absolute Error (MAE) compared to the ground truth values.

To train the bits estimation model, four hundred UHD sequences (80% of the sequences) from the Video Complexity Dataset [16] are used as the training set, and the remaining (20%) is used as the validation set. The sequences are encoded at 24 fps using VVenC v1.7.0⁴ [3] with the *faster* preset.

⁴<https://github.com/fraunhoferhhi/vvenc>, last access: Feb 20, 2023.

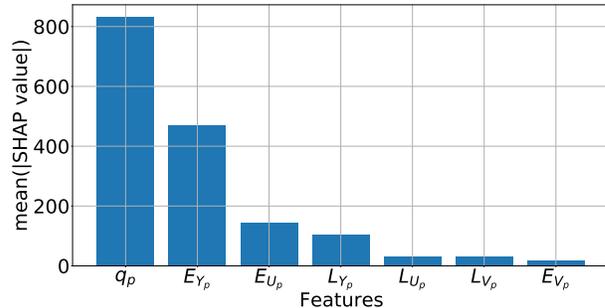


Fig. 3: Relative importance of features in bits prediction for I-frames.

Table 1: Bits estimator performance for various values of *max_depth*.

<i>max_depth</i>	MAE	R^2	Model Size
4	409.31 kb	0.81	0.22 Mb
8	217.25 kb	0.92	3.21 Mb
12	172.59 kb	0.93	31.34 Mb
16	155.20 kb	0.93	138.87 Mb

Results: Fig. 3 shows the relative importance of the considered features in estimating bits of I-frames using SHAP values [17]. It is observed that q_p is the most important feature, followed by the E_{Y_p} feature. Table 1 analyzes the model size and the prediction accuracy for I-frames for various values of *max_depth*. When *max_depth* = 12, MAE is observed to be 172.59 kb, while R^2 and model size are 0.93 and 31.34 Mb, respectively. When *max_depth* = 16, MAE is reduced to 155.20 kb, while R^2 and model size are 0.93 and 139.87 Mb, respectively. Since further increasing *max_depth* does not improve the results, *max_depth* = 12 is used in the following experiments. The scatter-plot in Fig. 4 depicts the correlation between the ground truth and model predictions of b for the considered values of *max_depth*. A strong correlation between the predictions and ground truth is observed when *max_depth* is set as 12.

3.2. Rate Control Performance

Experimental Setup: To evaluate the performance of the proposed method for the first pass of rate control [5], the VCA and bits estimator components are executed offline, and the output of the bits estimator part is used for the second pass in VVenC. The described method is combined with VVenC v1.7.0⁴ [3]. The RD performance is evaluated using the all-intra configuration at *faster* preset [4], without temporal subsampling. A predefined target rate was obtained from CTC-like coding with a fixed QP for each test. The presented coding efficiency was measured in Bjøntegaard Delta (BD) rate differences using JVET's CTC sequences for SDR classes A1 and A2 [18], and Fraunhofer HHI's public Berlin test set [19]. Following the requirements of the proposed

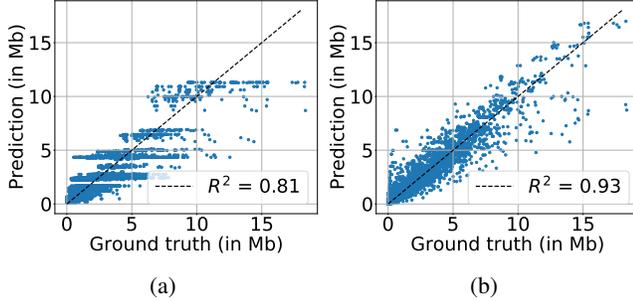


Fig. 4: Ground truth bits versus predicted bits per I-frame for max_depth values (a) 4, and (b) 12.

method, the test set is converted to 8-bit and 30 fps sequences. All evaluations are performed on an Intel Xeon E5-2697A v4 cluster with Linux OS and a GCC 7.3.1 compiler using eight CPU threads. The combined YUV BD-rate is calculated based on a weighted PSNR sum across all components [20]:

$$PSNR_{YUV} = \left(\frac{6 PSNR_Y + PSNR_U + PSNR_V}{8} \right). \quad (5)$$

The anchor is VVenC encoding of the test sequences using a fixed-QP all-intra setting and *faster* preset. The time required for the VCA and bits estimator steps was validated compared to the first pass of VVenC. The experiment was performed single-threaded on the same simulation platform.

Results: In the first experiment, the runtime of the methods is compared. The analysis has shown that the first pass speed of 2pRC is 0.40 frames per second (fps), while the proposed method yields over 40 fps. Hence, the first pass of the proposed method is a hundred times faster than the first pass of 2pRC. The overall encoding time for the end-to-end application can be found in Fig 5. The encoding time of the proposed scheme is 32.17% lower than 2pRC and 3% lower than fixed QP encoding.

The second experiment compares the rate-distortion (RD) performance of the methods. The results vary across the test sequences (*cf.* Table 2); on average, the proposed scheme results in a 2% BD-rate increase compared to VVenC’s original 2pRC method. Hence, the proposed method achieves an efficiency close to the fixed QP reference and 2pRC of VVenC.

The third experiment analyzes the bitrate deviation. It is observed that the bitrate deviation is close to zero for all classes, indicating that the proposed method does not deteriorate the bitrate accuracy. The behavior is similar to the 2pRC in VVenC.

Additionally, the performance of the proposed method was validated for significant changes in video content by testing a combined video sequence consisting of a concatenation of the eight Berlin sequences [19]. The results in Table 2 for the BerlinMix4K sequence show an accuracy roughly comparable to the average of other sequences. The worst-case scenario is also reviewed to systematize the results and evaluate the operating points. To achieve the worst-case, gaussian

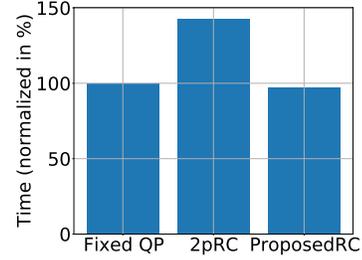


Fig. 5: Comparison of the overall encoding time using the considered rate control methods.

Table 2: Results compared to the fixed-QP encoding using 2pRC and the proposed method (*cf.* Section 2).

Dataset	Sequence	2pRC BD_{YUV} [%]	Proposed RC BD_{YUV} [%]	Noise BD_{YUV} [%]
JVET UHD	Tango1	0.09	8.35	33.55
	FoodMarket4	1.68	-0.64	17.98
	Campfire	0.17	0.31	24.30
	CatRobot	0.18	0.81	22.94
	DaylightRoad2	0.21	4.96	28.48
	ParkRunning3	0.02	-1.15	13.83
JVET UHD	Average	0.39	2.11	23.51
Berlin set	BerlinCrossroadsCrop4K	0.17	0.44	27.57
	ChestnutTreeCrop4K	0.26	0.34	18.69
	March18thSquareCrop4K	0.11	7.25	15.62
	NeptuneFountainCrop4K	0.93	2.67	19.09
	OberbaumCrop4K	0.18	0.29	17.24
	QuadrigaCrop4K	0.22	0.30	28.77
	ReichstagIntoTreeCrop4K	0.38	2.57	18.04
	SpreeCrop4K	1.23	2.15	23.40
Berlin set	Average	0.44	2.00	21.05
	BerlinMix4K	0.06	2.69	14.38

The sequences were resampled to 2160p 8bit 30fps.

white noise $\hat{b}_p = \mathcal{N}(\mu, \sigma^2)$, with mean μ and standard deviation σ equal to the target per frame bits, simulated as the input to the second pass of the 2pRC encoding. The confrontation of rate control with the arbitrary noise signal allows us to validate the new approach and determine the limit of possible deterioration. The proposed rate control method achieves, on average, significantly lower compression losses ranging from slight gains of -1.15% to losses up to 8.35% , compared to noise with the degradation above 20% on average and ranging up 33.55% .

4. CONCLUSIONS

A simplified first pass operation for all-intra 2pRC in VVenC, a practical *Versatile Video Coding* encoder, has been presented. Around the preset *faster*, the proposed method reduces the encoding time by 32.17%, on average causing a 2% BD-rate increase over the default 2pRC method, *i.e.*, requiring only 2% more bits to produce the same objective quality. The proposed first pass is realized using the Video Complexity Analyzer (VCA) and an RF model to predict the required per-frame bits from the VCA features. This pre-analysis is used instead of a complete encoding with a reduced toolset used in VVenC. Especially for the preset *faster*, the approach yields significant time savings of 32%, achieving runtime on par with that of fixed QP encoding.

5. REFERENCES

- [1] B. Bross *et al.*, “Overview of the Versatile Video Coding (VVC) Standard and its Applications,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3736–3764, 2021.
- [2] G. J. Sullivan *et al.*, “Overview of the High Efficiency Video Coding (HEVC) Standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [3] A. Wieckowski *et al.*, “VVenC: An Open And Optimized VVC Encoder Implementation,” in *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2021, pp. 1–2.
- [4] J. Brandenburg *et al.*, “Pareto-optimized coding configurations for VVenC, a fast and efficient VVC encoder,” in *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*, 2021, pp. 1–6.
- [5] C. R. Helmrich *et al.*, “Visually Optimized Two-Pass Rate Control for Video Coding Using the Low-Complexity XPSNR Model,” in *2021 International Conference on Visual Communications and Image Processing (VCIP)*, 2021, pp. 1–5.
- [6] C. R. Helmrich *et al.*, “A Scene Change and Noise Aware Rate Control Method for VVenC, An Open VVC Encoder Implementation,” in *2022 Picture Coding Symposium (PCS)*, 2022, pp. 241–245.
- [7] J.-R. Ohm *et al.*, “Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC),” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1669–1684, 2012.
- [8] A. Wieckowski *et al.*, “VVenC: an open optimized VVC encoder in versatile application scenarios,” in *Applications of Digital Image Processing XLIV*, A. G. Tescher and T. Ebrahimi, Eds., vol. 11842, International Society for Optics and Photonics. SPIE, 2021, p. 118420H.
- [9] V. V. Menon *et al.*, “Green Video Complexity Analysis for Efficient Encoding in Adaptive Video Streaming,” in *First International ACM Green Multimedia Systems Workshop (GMSys '23)*, 2023.
- [10] V. V. Menon *et al.*, “Content-adaptive encoder preset prediction for adaptive live streaming,” in *2022 Picture Coding Symposium (PCS)*, 2022, pp. 253–257.
- [11] V. V. Menon *et al.*, “JND-aware Two-pass Per-title Encoding Scheme for Adaptive Live Streaming,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.
- [12] V. V. Menon *et al.*, “Transcoding quality prediction for adaptive video streaming,” in *Proceedings of the 2nd Mile-High Video Conference*, 2023, p. 103–109.
- [13] V. V. Menon, “Video Coding Enhancements for HTTP Adaptive Streaming,” in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, p. 6905–6909.
- [14] ITU-T, “P.910 : Subjective video quality assessment methods for multimedia applications,” Nov. 2021.
- [15] C. R. Helmrich *et al.*, “Finalization of VVenC’s Screen Content Detector and Two-Pass Rate Control Using Pre-Filtering Statistics,” *submitted to IEEE International Conference on Image Processing (ICIP)*, 2023.
- [16] H. Amirpour *et al.*, “VCD: Video Complexity Dataset,” in *Proceedings of the 13th ACM Multimedia Systems Conference*, 2022.
- [17] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” in *Advances in Neural Information Processing Systems 30*, 2017, pp. 4765–4774.
- [18] F. Bossen *et al.*, “VTM common test conditions and software reference configurations for SDR video,” in *20th Joint Video Experts Team on Video Coding Meeting, Virtual*, Oct. 2020.
- [19] B. Bross *et al.*, “AHG4 Multiformat Berlin Test Sequences,” in *JVET-Q0791*, 2020.
- [20] I.-T. HSTP-VID-WPOM and I. T. 23002-8, “Working practices using objective metrics for evaluation of video coding efficiency experiments,” 2021. [Online]. Available: <https://www.itu.int/pub/T-TUT-ASC-2020-HSTP1>