A SCENE CHANGE AND NOISE AWARE RATE CONTROL METHOD FOR VVENC, AN OPEN VVC ENCODER IMPLEMENTATION

Christian Helmrich, Christian Bartnik, Jens Brandenburg, Valeri George, Tobias Hinz, Christian Lehmann, Ivan Zupancic, **Adam Wieckowski**, Benjamin Bross, and Detlev Marpe

a Presentation at the Picture Coding Symposium 2022

Fraunhofer HHI

A Scene Change and Noise Aware Rate Control Method for VVenC

Outline

1. Introduction, Motivation
2. Frame Type Adaptation (FTA) for Key Frames
3. GOP-Wise Two-Pass Rate Control (RC) Design
4. Noise Level Estimation and Limitation of QPs
5. Performance Evaluation
6. Summary and Conclusion

1. Introduction, Motivation

Rate Control and VVenC

- **Rate control** (RC) methods are inevitable in video coding applications like streaming
  - Control of average (= mean) coding bit-rates across full videos, or scenes thereof
  - Control of bit rate variance across scenes, or groups of pictures (GOPs), of videos
- **VVenC:** open* Versatile Video Coding (VVC) compliant video encoder [ICMEW, 2021]
  - Featuring a recently developed **two-pass RC** method [Helmrich *et al*., VCIP 2021]
  - Uses *R-QP* (instead of *R*-λ) model with two encoding passes → sequence-wise RC
- VVenC's two-pass RC initially designed for **file** based operation, using short I-periods

* Code: https://github.com/fraunhoferhhi/vvenc

1. Introduction, Motivation

Issues with Existing Methods

- **Issue 1:** both short and long input independent I-periods have specific disadvantages
  - Frequent I-frames consume a lot of the coding bits available in small bit budgets
  - Long I-periods lead to diminishing efficiency returns, may lead to **visual artifacts**
- **Issue 2:** Sequence-wise RC operation doesn't work with 'on-the-fly' **stream** encoding
  - Live or pipe based encoding: every GOP is passed to encoder, one after the other
- **Issue 3:** strongly varying picture statistics between short scenes often destabilize RCs
  - Most prominent case: varying sensor noise or film grain levels (low light, credits)

Fraunhofer HHI

1. Introduction, Motivation

Contributions of This Paper

- **Solution 1:** input adaptive key frame type selection → frame type adaptation (FTA)
  - For reliable use of **long I-periods** without risk of visual artifacts or efficiency loss
  - Reaches the advantages of both short and long I-periods on a GOP-by-GOP basis
- **Solution 2:** extension of sequence-wise RC design to allow for GOP-wise operation
  - GOP-wise two-pass RC with **sliding window** → usable with 'on-the-fly' encoding
- **Solution 3:** 'minimum statistics' based noise level estimation with block QP limiting
  - Prevents excessive rate consumptions in 2^{nd} encoding pass → **stabilization** of RC.

A Scene Change and Noise Aware Rate Control Method for VVenC

Outline

1. Introduction, Motivation
2. Frame Type Adaptation (FTA) for Key Frames
3. GOP-Wise Two-Pass Rate Control (RC) Design
4. Noise Level Estimation and Limitation of QPs
5. Performance Evaluation
6. Summary and Conclusion

2. Frame Type Adaptation (FTA)

Background, Overview

- Since H.264/AVC scalable coding, use of usually dyadic hierarchical coding structure
  - Each coded picture is assigned a **temporal level** *TL*, coding order ≠ display order
  - All pictures belong to one specific group of pictures (**GOP**), e.g. GOP size G = 32
- Types of TL = 0 frames: **Intra-only** (I) for random-access point, **Inter** (B, P) otherwise
  - **FTA**: use long I-periods with few fixed I-frames, add I-frame near scene changes
  - **Reasoning**: *TL*-0 B or P frames near scene changes use many bits, type I is better

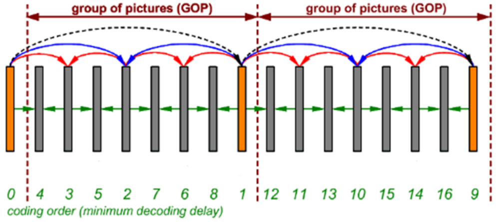
RA coding: I … B … B … I … B … B … I … B …

with FTA: I … B … B … I … B … I … I … B …

TL-0 frames are equivalent to key frames herein

scene change detected in **this GOP**

orange: $TL = 0$
gray: $TL > 0$
$I_G = I / G = 3$



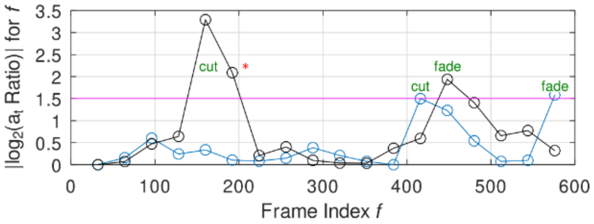
2. Frame Type Adaptation (FTA)

Implementation Details

- **Assumption:** I-period is integer multiple of GOP size G, first adapted frame at $f_P > G$
  - No FTA is needed when first TL-0 frame after a scene change is a regular I frame
- **Detection process:** determine spatio-temporal visual activity a_t at each TL-0 frame f_P
  - a_t defined as a_k in XPSNR metric [Helmrich *et al.*, JITU '20], with two differences:
  - **Picture-wise**, not block-wise a_t, last picture is **last *TL*-0**, not last displayed picture
  - Switch TL-0 frame to type I if* $a_t(f_P) > T \cdot a_t(f_M)$ or $T \cdot a_t(f_P) < a_t(f_M)$, with $f_M = f_P - G$

| scene change

Top: – MarketPlace, bottom: – RitualDance, right: – threshold $T = 2^{1.5}$



*: detections in successive key frames forbidden

A Scene Change and Noise Aware Rate Control Method for VVenC

Outline

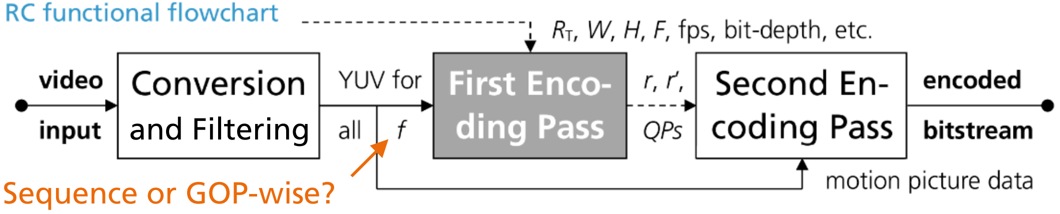
1. Introduction, Motivation
2. Frame Type Adaptation (FTA) for Key Frames
3. GOP-Wise Two-Pass Rate Control (RC) Design
4. Noise Level Estimation and Limitation of QPs
5. Performance Evaluation
6. Summary and Conclusion

3. GOP-wise Two-pass Rate Control

Background, Overview

- **First RC pass:** fast R-D optimized encoding with fixed quantization parameters (QPs)
  - Like non-RC coding; overall base QP estimated from video size, target rate R_{target}
  - Goal: collect frame-wise QP, λ, rate across ~~entire video~~ → **short video look-ahead**
- **Second RC pass:** final encoding with adapted (based on first-pass statistics) CTU QPs
  - QPs adapted for visual quality & to match R_{target} across ~~video~~ → **analysis window**
  - Here, look-ahead size set to 1 G, analysis window $\boldsymbol{A}$ to min(8 G, I-period/G) + 1 G

RC functional flowchart



sequence-wise RC: **f** means all frames in video

GOP-wise RC: **f** means all frames in **next GOP**

(i.e., frames in the lookahead buffer)

Fraunhofer HHI

3. GOP-wise Two-pass Rate Control

Implementation Details

- **First RC pass:** read and encode a new GOP, move analysis window *A* forward in time
  - With statistics of previously coded I-period, derive 2nd-pass QPs for the new GOP

- **Second RC pass:** estimate base QP (for e. g. filtering) from video size, I-period, R_{target}
  - Final bit allocation based on instantaneous 1st-pass rates → R_{target} in **all** I-periods*
  - Additional 2nd-pass **QP constraints** for improved stability across scenes (Sec. III.B)
  - Forbid mean frame QPs from changing too abruptly over time and between *TL*s.

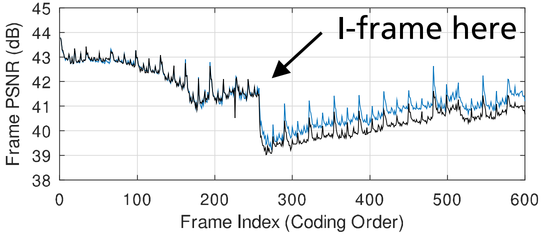
intermediate QP: $QP'_{\text{base}} = \text{round}\left(40 - 1.5D_1\sqrt{R_{\text{target}}/500000} - 0.5\log_2 I_G\right)$

2nd-pass base QP: $QP''_{\text{base}} = \text{round}\left(QP'_{\text{base}} + c_{\text{high}} \cdot \max(0; 24 - QP'_{\text{base}})\right)$

D_1 and c_{high} depend on video size

* sequence-wise RC: only on average across video

– effect of QP constraints:



A Scene Change and Noise Aware Rate Control Method for VVenC

Outline

1. Introduction, Motivation
2. Frame Type Adaptation (FTA) for Key Frames
3. GOP-Wise Two-Pass Rate Control (RC) Design
4. Noise Level Estimation and Limitation of QPs
5. Performance Evaluation
6. Summary and Conclusion

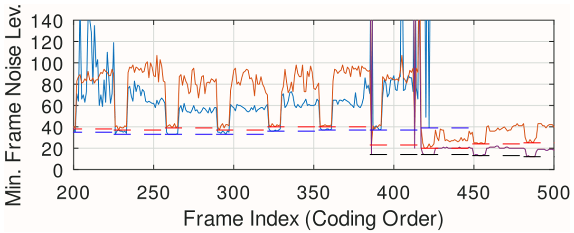
4. Noise Level Estimator, QP Limiter

Background, Overview

- **Theory:** fine quantization of picture noise visually irrelevant, costly, destabilizes RC
  - Some actually prefer denoising effect of encoder, or utilize **film grain synthesis**
  - Noise isn't predictable → high coding bit demand **better spent on other details**

- **Approach:** find picture noise level in 1st pass, apply proportional QP limit in 2nd pass
  - Simple noise level estimator: **minimum statistics** based estimator [EuroSp 1993]
  - Convert noise level into step size, find nearest corresp. QP, take it as **lower limit**

Noise level estimates in the MarketPlace video: Minima across all frame minima in every GOP, separately for – low, – mid, – high luminance



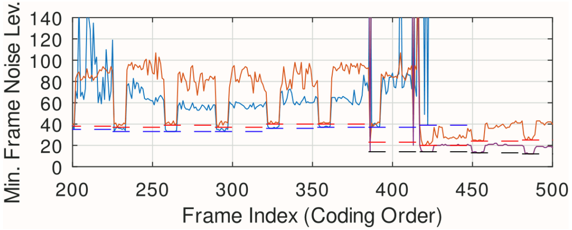
4. Noise Level Estimator, QP Limiter

Implementation Details

- **First RC pass:** estimate noise level per CTU block *k* in 8 different luma intensity regions
  - Input: min. visual activity (see a_k) and, since VVenC 1.7, min. MCTPF motion error*

- **Second RC pass:** clip QP before encoding a CTU, reset estimator after encoding a GOP
  - Again depending on luma average in CTU (one of 8 regions → details in Sec. IV.A)
  - When bits are saved due to the limiting, RC can try to spend them **elsewhere later**
  - Example: bits saved in noisy scene spent in following low-noise fast moving scene.

Noise level estimates in the MarketPlace video: Minima across all frame minima in every GOP, separately for – low, – mid, – high luminance



*: use of MCTPF motion error not described in paper

A Scene Change and Noise Aware Rate Control Method for VVenC

Outline

1. Introduction, Motivation
2. Frame Type Adaptation (FTA) for Key Frames
3. GOP-Wise Two-Pass Rate Control (RC) Design
4. Noise Level Estimation and Limitation of QPs
5. Performance Evaluation
6. Summary and Conclusion

5. Performance Evaluation

RA - Left: fixed-QP (no RC), Right: 2-pass RC

- **VVenC 1.6**, Bjøntegaard delta (BD)-rates, JVET's common test conditions [JVET-Y2010]*
- **Test 1:** evaluation of FTA when going from I-period **1 sec** to **4 sec** → loss turned to gain
- **Test 2:** evaluation of GOP-wise RC, comparison to **VTM 14 + RC** [JVET-Y0105] → good ☺
  - R_{target} and BD-rate references: resp. VVenC or VTM fixed-QP CTC results (QPs 22–37)

Test 1

Resolution Class	Without FTA 1→ 4	With FTA 4→ 4	Overall 1→ FTA 4
UHD A½	–2.07%	0.00%	–2.07%
HD B	–1.88%	–1.93%	–3.95%
SD C	–4.46%	0.06%	–4.40%
Overall	–2.64%	–0.63%	–3.32%
MarketPl.	**8.24%**	**–9.27%**	**–1.74%**

Test 2

Resolution Class	VTM14.0 no QPA **1**	VVenC1.6 III.B off **4**	VVenC1.6 III.B on **4**	Sequence wise RC **4**
UHD A½	10.04%	2.40%	**0.65%**	–0.46%
HD B	5.45%	4.60%	**1.31%**	0.63%
SD C	4.02%	5.47%	**2.10%**	0.91%
Overall	6.91%	3.95%	**1.26%**	0.27%

numbers in table headers: I-periods

*: YUV averages; length of all class-A sequences set to 10 sec

A Scene Change and Noise Aware Rate Control Method for VVenC

Outline

6. Summary and Conclusion

Rate Control (RC) Encoding in VVenC

- This paper presented three recent enhancements to VVenC's rate control algorithm:
- **Frame type adaptation** improves encoding with long I-periods around scene changes
- **Changes to RC** enable GOP-wise 'on-the-fly' encoding of video streams, not only files
- **QP limiter** based on noise level estimation improves both RC stability & bit allocation
- Excellent performance of FTA and GOP-wise RC confirmed using BD-rate experiments
  - **Up to 9%** BD-rate gain due to FTA on JVET's CTC test set (**1.9% overall** on class B)
  - VVenC + GOP-wise RC outperforms JVET's VTM + RC [JVET-Y0105] by **about 5 – 8%**

- **Future work:** more RC stability and visual quality improvements via QP optimization.

Thank you for your attention!
The VVenC Team

→ https://github.com/fraunhoferhhi/vvenc

Annex – References

[ICMEW 2021] A. Wieckowski, J. Brandenburg, T. Hinz, C. Bartnik, V. George, G. Hege, C. Helmrich, A. Henkel, C. Lehmann, C. Stoffers, I. Zupancic, B. Bross, and D. Marpe, «**VVenC: An Open and Optimized VVC Encoder Implementation**», in *Proc. IEEE Int. Conf. Multim. & Expo Workshops (ICMEW)*, July 2021. https://ieeexplore.ieee.org/document/9455944/

[VCIP 2021] C. R. Helmrich, I. Zupancic, J. Brandenburg, V. George, A. Wieckowski, and B. Bross, «**Visually Optimized Two-Pass Rate Control for Video Coding Using the Low-Complexity XPSNR Model**», in *Proc. IEEE Int. Conf. on Visual Communications & Image Processing (VCIP)*, Munich, Dec. 2021. https://ieeexplore.ieee.org/document/9675364/

[JITU 2020] C. R. Helmrich, S. Bosse, H. Schwarz, D. Marpe, and T. Wiegand, «**A Study of the Extended Perceptually Weighted Peak Signal-to-Noise Ratio (XPSNR) for Video Compression with Different Resolutions and Bit Depths**», *ITU Journal: ICT Discoveries (Special Issue)*, vol. 3, no. 1, URL, May 2020. http://handle.itu.int/11.1002/pub/8153d78b-en

[EuroSp 1993] R. Martin, «**An Efficient Algorithm to Estimate the Instantaneous SNR of Speech Signals**», in *Proc. IEEE EuroSpeech*, Berlin, Sep. 1993. https://www.isca-speech.org/archive_v0/eurospeech_1993/e93_1093.html

[JVET-Y2010] F. Bossen *et al.*, «**VTM and HM common test conditions and software reference configurations for SDR 4:2:0 10-bit video**», ITU doc. *JVET-Y2010*, Feb. 2022. www.jvet-experts.org -> All meetings -> Number 25

[JVET-Y0105] G. Ren, J. Jia, J. Wang, and Z. Chen, «**AHG10: An improved rate control scheme**», ITU doc. *JVET-Y0105*, Jan. 2022; **see also:** ITU doc. *JVET-T0062*, Oct. 2020. www.jvet-experts.org -> All meetings -> Number 25