Note of A Scene Change and Noise Aware Rate Control Method for VVenC An Open VVC Encoder Implementation

Ushio Sorakado*

***Abstract*—This document is the note of the paper named: "Note of A Scene Change and Noise Aware Rate Control Method for VVenC An Open VVC Encoder Implementation", which has been accepted at Picture Coding Symposium (PCS), 2022. We access the paper via the following link: https://ieeexplore.ieee.org/document/10018041. The paper propose three extensions to the two-pass rate control (RC) adopted by VVenC: first, a frame type adaptation method is proposed to deal with scene cut; second, sequence-wise two-pass RC is modified to group-of-picture (gop)-wise lookahead based two-pass RC; third, a block based low-complexity noise level estimation is implemented to avoid large variations bit consumption caused by camera noise or film grain. The proposed method is implemented on VVenC-1.6.0 [1]. Experimental results show that the proposed method only sacrifices roughly 1% Bjøntegaard Delta-rate (BD-rate), however, enables two-pass RC could be used in on-the-fly scenarios.**

***Index Terms*—Two-pass rate control (RC), Visual quality assessment (VQA), lookahead.**

I. Introduction

The motivation of [2] is making two-pass RC scheme of [3] allow for on-the-fly operation and improving its performance in such scenarios, e.g. smotion picture content comprising short scenes of strongly varying statistics.

The contribution of [2] is summarized as follows.

Theoretical support of the method:
The theoretical support of the propsoed method is the same as [3]. Please refer to Note_of_Visually_Optimized_Two-Pass_Rate_Control_for_Video_Coding_Using_the_Low-Complexity_XPSNR_Model for detail.

Frame Type Adaptation for Key Frames:
The proposed frame type adaptation (FTA) algorithm modified the frame, whose temporal levels $l \leq 1$, from P to I, based on the ratio of frame-level spatio-temporal visual activity measure $a(f)$ [4] of current frame and correspoding frame.

Assuming that the Intra-only period I is an integer multiple of the group-of-picture (GOP) size G, every I_G^{th} key frame, with $I_G = I/G$, will represent an I-frame, starting at the first encoded frame (here, at frame index $f_0 = 0$). Let $f_P > G$ denote the index of the P-frame subjected to FTA detection and $f_M = f_P – G$ the index of the reference frame for f_P f, and both f_P and f_M pointing to key frames. Frame-level spatio-temporal visual activity $a(f)$ for the luma picture at every f may be derived, particularly for both f_P and f_M:

$$a_t(f) = max\left(a_{min}^2; \left(\frac{1}{4WH}\sum |h_s[x,y]| + |h_f[x,y]|\right)^2\right), \tag{1}$$

where W and H are the video width and height. The definitions of a_{min}^2, h_s are provided in [4] and omitted here for reasons of brevity. h_f represents a temporal high-pass that, is formulated as $h_f[x,y] = 2\left(s_f[x,y] - s_{f-G}[x,y]\right)$, where s holds the luma picture samples at the given frame index and x, y are the horizontal, vertical sample indices, respectively. Finally, the ration of $a_t(f_P)$ and $a_t(f_M)$ is calculated and compared with a pre-defined threshold T, where T is set as 1.5. If

$$|log_2(a_t(f_P)/a_t(f_M)| > T, \tag{2}$$

f_P is determined as I frame.

GOP-Wise Two-Pass RC:
As sequence-wise two-pass RC proposed by [3] need to encode the whold sequence in each pass RC, which is not possible with live sources in which the video sequence is not available to the encoder in its entirety a priori. Thus, a GOP-wise lookahead baed two-pass rate control is proposed in this paper. Hereafter, only the modifications to sequence-wise two-pass RC is introduced, please refer to Note_of_Visually_Optimized_Two-Pass_Rate_Control_for_Video_Coding_Using_the_Low-Complexity_XPSNR_Model if you don't understand about it [3].

Let A enote the number of frames for which data is collected in each pass RC, in other words, the size of frame group in each pass RC. $A = min(8G; I) + G$, the statistics of the last Intra period (limited to at most 8G) of previously encoded frames are added to the GOP of lookahead statistics, and with typical $I \leq 8G$, the sliding window will cover $I_G + 1$ GOPs.

Since no sequence-mean first-pass data are available when setting QP''_{base} at start of the second RC pass. The formulation of QP''_{base} is modified as

$$QP''_{base} = round\left(QP'_{base} + c_{high} \cdot max\left(0; 24 - QP'_{base}\right)\right), \tag{3}$$

where

$$QP'_{base} = round\left(40 - 1.5D_1\sqrt{R_{target}/500000} - 0.5log_2 I_G\right). \tag{4}$$

where $D_1 = \sqrt{\frac{3840 \cdot 2160}{W \cdot H}}$. The other process is the same as [3].

To avoid destabilization of the RC's operation occurring when the features change abruptly upon scene changes, the frame-level QP change constraints are refined and extended as follows.

1) QP''_f at level l shall lie in range $[QP''_{curL} \pm c_{curL}]$, where QP''_{curL} is the QP of the last coded frame at the same l; $c_{curL} = 5 + I_G$ during a scene change and $c_{curL} = max(3; 6 - [\frac{l}{2}]$ otherwise.
2) QP''_f at level l shall be larger than QP''_{prevL}, the QP of the last coded frame one level below l (i.e., at level l–1), with $l > 1$.
3) QP''_f at level $l \leq 1$ shall lie in range $[1 + QP_{avg/2}, QP_{max}]$, with QP_{avg} holding the mean of all past QP''_f in P and $QP_{max} = 63$.
4) QP''_f at level l shall lie in range $[l + QP''_{base/2}, QP_{max}]$, with the second-pass base QP estimate QP''_{base}.

Noise Level Estimation And Limiting Of QPs:
For detail please refer to [2].

References

[1] Adam Wieckowski, Jens Brandenburg, Tobias Hinz, Christian Bartnik, Valeri George, Gabriel Hege, Christian Helmrich, Anastasia Henkel, Christian Lehmann, Christian Stoffers, Ivan Zupancic, Benjamin Bross, and Detlev Marpe, "Vvenc: An open and optimized vvc encoder implementation," in *Proc. IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, 2021, pp. 1–2.

[2] Christian R Helmrich, Christian Bartnik, Jens Brandenburg, Valeri George, Tobias Hinz, Christian Lehmann, Ivan Zupancic, Adam Wieckowski, Benjamin Bross, and Detlev Marpe, "A scene change and noise aware rate control method for vvenc, an open vvc encoder implementation," in *2022 Picture Coding Symposium (PCS)*. IEEE, 2022, pp. 241–245.

[3] Christian R Helmrich, Ivan Zupancic, Jens Brandenburg, Valeri George, Adam Wieckowski, and Benjamin Bross, "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)*. IEEE, 2021, pp. 1–5.

[4] Christian R Helmrich, Sebastian Bosse, Heiko Schwarz, Detlev Marpe, and Thomas 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*, vol. 3, no. 1, pp. 65–72, 2020.