SSIM Motivated Quality Control for Versatile Video Coding

Meng Wang*, Shiqi Wang*, Junru Li†, Li Zhang‡, Yue Wang§ and Siwei Ma†
* City University of Hong Kong, Hong Kong, China.
† Peking University, Beijing, China.
‡ Bytedance Inc., San Diego CA., USA.
§ Bytedance (HK) Limited., Hong Kong, China.

***Abstract*—In this paper, we investigate the encoder control algorithm driven by the visual quality in Versatile Video Coding (VVC) standard. The proposed framework shifts the traditional bit rate-centered paradigm to the quality-centered paradigm, and adopts the repeatedly proven perceptual quality measure structural similarity index (SSIM) as the core in guiding the behaviors of encoder. In particular, a reduced-reference evaluation metric which could statistically establish the relationship between the SSIM and quantization parameter, is adopted for quality control. Extensive experiments verify the effectiveness of the proposed method, leading to precise quality control in the VVC Test Model (VTM-9.0).**

I. Introduction

Recent years have witnessed the exponential growth of video-oriented services fueled by the proliferation of video data and acquisition devices. The gigantic scale of video data poses new challenges to video compression and transmission. Versatile Video Coding (VVC) standard [1] is an emerging video coding standard developed by Video Coding Experts Group (VCEG) and Moving Picture Experts Group (MPEG), aiming at further promoting the compression efficiency compared to the previous High Efficiency Video Coding (HEVC) [2] standard. Numerous advanced coding tools have been thoroughly investigated and adopted during the standardization process, including Multi-Tree Types (MTT) partitioning [3], enhanced inter and intra prediction [4], [5], [6], multiple transform selection [7], dependent quantization [8] and adaptive loop filter [1]. As such, it has been reported that VVC achieves 25% to 36% compression performance improvement compared with HEVC standard [9].

It has been widely acknowledged that rate control plays a central role in encoder optimization, in particular when many digital video applications are constrained by limited bandwidth or storage space. However, with the advances of the network technology such as 5G and 6G, the growth of bandwidth and low cost of storage make rate control less essential, as bit rate might be no-longer regarded as the critical bottleneck in warranting the high quality video services [10]. Recently, many recent studies acknowledge the importance of quality of experience as the ultimate receiver of videos is the human visual system, and it has been widely accepted that it is difficult to equate bitrate to the visual quality due to the fact that video content influences the visual quality and bit rate in different ways [11], [12]. As such, it is highly desired to shift the traditional bitrate centralized to quality centralized encoder optimization paradigm, and leverage the perceptual quality to define the level of services in the codec implementation of the state-of-the-art video coding standard.

The structural similarity (SSIM) index [13] has been broadly accepted as a well-behaved quality measure due to the excellent trade-off between the complexity and accuracy, such that it has been widely employed in video coding optimization such as rate control [14], rate-distortion modeling [15], and coding parameter adjustment [16]. In this paper, we make an attempt to achieve quality control based on SSIM in VVC, in an effort to encode the video that matches the target quality instead of the bitrate. Furthermore, we impose the smoothness quality constraint to ensure that the encoded video has constant quality without large fluctuation in terms of the quality measure. More specifically, the proposed scheme is based on a reduced-reference quality assessment metric that mimics the divisive normalization principle behind the SSIM index in Discrete Cosine Transform (DCT) domain. Moreover, based on the constantly capturing of the characteristics of each frame, given the target quality the optimal coding parameter can be inferred based on the distortion-quantization model. Experimental results show that the proposed method is effective in precisely achieving the target quality level with relatively smooth quality variations.

II. The Proposed Quality Control Algorithm

SSIM has been widely recognized to be effective in capturing the supra-threshold coding artifacts [17]. As such, given the target SSIM quality level $SSIM_t$, we attempt to achieve the target quality by imposing the smoothness on the quality for individual frame with the manipulation of the quantization parameter QP, which can be formulated as follows,

$$QP^{(p)} = \arg\min_{QP} |\mathcal{G}(QP) - SSIM_t|, \tag{1}$$

where $QP^{(p)}$ is the optimal quantization parameters for the p-frame encoding. $\mathcal{G}(QP)$ characterizes the relationship between QP and the perceptual quality implied by SSIM. To achieve the quality control for video coding, it is necessary to establish the relationship between the coding parameters and perceptual quality of encoded videos based on the content characteristics,

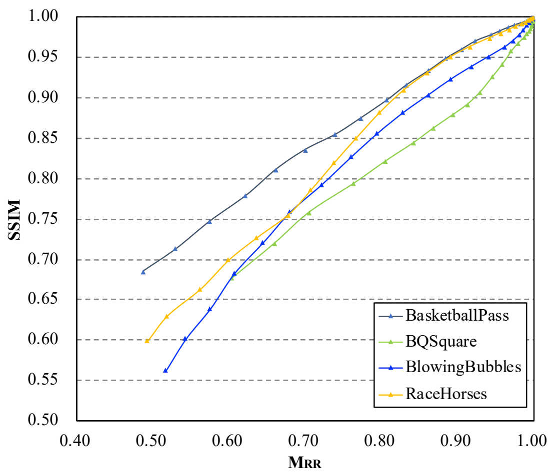

Fig. 1. Illustration of the relationship between M_{RR} and SSIM under low delay P configuration with varied QPs in VVC.

such that the optimal QP can be inferred before coding the current frame.

A straightforward way to instantiate $\mathcal{G}(QP)$ is to explore the relationship between the SSIM of encoded videos and associated QP. However, due to the fact that SSIM is a full-reference measure which requires both the original and the distorted frames for further calculation, in practice, the distorted frames are not accessible unless they have been fully reconstructed. Therefore, SSIM cannot be directly applied in the quality control framework, and a reduced-reference SSIM (RR-SSIM) is highly desirable in such scenario. In [18], an RR-SSIM was investigated to estimate the actual SSIM based on a multi-scale and multi-orientation divisive normalization transform, which achieves high accuracy along with high computational complexity. In the proposed quality control algorithm, we employ a low-complexity RR-SSIM [15] based on the features from DCT domain to effectively speculate the perceived quality of encoded videos within the coding loop. In particular, the full-reference SSIM index in DCT domain is formulated as follows [19],

$$SSIM(\mathbf{c},\hat{\mathbf{c}}) = \left\{1 - \frac{(C(0)-\hat{C}(0))^2}{C(0)^2+\hat{C}(0)^2+N\cdot z_1}\right\}\times \left\{1-\frac{\sum_{i=1}^{N-1}(C(i)-\hat{C}(i))^2}{\sum_{i=1}^{N-1}(C(i)^2+\hat{C}(i)^2)+N\cdot z_2}\right\}, \quad (2)$$

where $C(i)$ and $\hat{C}(i)$ denote the original and distorted DCT coefficients of signal $\mathbf{c}$ and $\hat{\mathbf{c}}$, respectively. N represents the total number of samples. z_1 and z_2 are two factors related to bit-depth to avoid that the denominators are too close to zero.

Since the ultimate goal is to regulate the quality of encoded videos given the target SSIM quality level, an encoder-friendly reduced-reference model is highly desirable which could exhibit one-to-one mapping property with the full-reference SSIM and enjoy the convenience of calculation with the coding loop. Based on our previous work, a reduced-reference format of SSIM index can be represented with M_{RR} [15], which is inspired by the design philosophy of the divisive normalization behind the DCT-domain SSIM index. In particular, M_{RR} can be described as [15],

$$M_{RR} = \left(1-\frac{D_0}{2\sigma_0^2+x_1}\right)\left(1-\frac{1}{K-1}\sum_{k=1}^{K-1}\frac{D_k}{2\sigma_k^2+x_2}\right), \quad (3)$$

where D_k denotes the distortions in terms of the mean square error (MSE) at the k-th sub-band. σ_k represents the variance of the transform coefficients at the k-th sub-band.

To establish such an encoder-friendly RR-SSIM model in terms of the M_{RR}, pre-searching is conducted before coding each frame to acquire the statistical information, which further serves as the guidance of quality control. More specifically, each frame is firstly divided into 8×8 non-overlapping blocks, and then genuine intra and inter predictions are delicately traversed. It should be noted that the pre-searching will not significantly boost the computational overhead to the encoder due to the fixed block size. Subsequently, after obtaining the pre-searched residual signals, DCT with the dimension of 4×4 is repeatedly applied to each sub-area of the 8×8 blocks. We gather the transform coefficients within the same frequency sub-band for distribution modelling. As such, the transform coefficients from 16 sub-bands could yield 16 distribution models, representing the characteristics of the frequency-domain information in different sub-bands.

Furthermore, to estimate D_k, we model the pre-searched transform coefficients with Laplacian distribution, which is a classical distribution model repetitively employed in video coding task owing to the excellent trade-off between model accuracy and computational complexity in model parameter derivation. As such, given the transform coefficients in the k-th sub-band, the distribution model can be formulated as,

$$f(x) = \frac{\Lambda_k}{2}\cdot e^{-\Lambda_k\cdot|x|}. \quad (4)$$

where Λ_k is the distribution parameter that can be derived with the variance v_k regarding the transform coefficients of the pre-searched residuals,

$$\Lambda_k = \frac{\sqrt{2}}{v_k}. \quad (5)$$

In this manner, the D_k can be estimated with the quantization step size Q as [20],

$$\begin{aligned} D_k &= \int_{-Q(1-\gamma)}^{Q(1-\gamma)} x_k^2 f(x_k) dx_k \\ &+ 2\sum_{n=1}^{\infty} \int_{Q(n-\gamma)}^{Q(n+1-\gamma)} (x_k - nQ)^2 f(x_k) dx_k \\ &= \frac{\eta_k \cdot e^{\eta_k \gamma}[2 + \eta_k(1-2\gamma)] + 2 - 2e^{\eta_k}}{\Lambda_k^2 (1 - e^{\eta_k})}, \end{aligned} \tag{6}$$

where

$$\eta_k = \Lambda_k \cdot Q. \tag{7}$$

Here, γ is the quantization rounding offset, which equals to $\frac{1}{3}$ and $\frac{1}{6}$ for I slice and P/B slice [21], respectively. Moreover, Q can be directly converted from the quantization parameter QP [1].

We show the relationship between M_{RR} and actual SSIM for four test sequences coded by VTM-9.0 [22], as illustrated in Fig. 1. The M_{RR} and actual SSIM are averaged with respect to frames. The base QP varies from 2 to 50 with the interval of 2, with the goal of covering wide-ranging encoding quality levels. It can be noticed that the SSIM and M_{RR} hold approximate linear relationship, and the slope highly relies on video content. Therefore, it is applicable to employ M_{RR} criteria to simulate the SSIM index in quality-constraint coding. Given the target perceptual quality of a video, which corresponds to the target SSIM score in our method, pursuing the target SSIM is converted to achieve the destined M_{RR} with,

$$TM_{RR}^{(p)} = a^{(p)} \cdot SSIM_t^{(p)} + b^{(p)}, \tag{8}$$

where $TM_{RR}^{(p)}$ is the target M_{RR} value for the p-th frame. $a^{(p)}$ and $b^{(p)}$ denote the slope and the intercept of the linear model regarding the p-th frame, which can be derived and updated on-the-fly based on video content. $SSIM_t^{(p)}$ represents the target SSIM. Furthermore, by integrating the D-Q relationship in Eqn. (6) into Eqn. (3), M_{RR} can be employed to instantiate the $\mathcal{G}(QP)$ as,

$$\mathcal{G}(QP) = M_{RR}. \tag{9}$$

To achieve the quality control, we attempt to minimize the gap between the target quality $TM_{RR}^{(p)}$ and the estimated quality $\mathcal{G}(QP)$ by exhaustively traversing different QP as follows.

$$QP^{(p)} = \arg\min_{QP} |\mathcal{G}(QP) - TM_{RR}^{(p)}|. \tag{10}$$

In this manner, $QP^{(p)}$ is regarded as the optimal quantization parameter for the p-th frame coding.

For the convenience of parameters updating, we regard $(1,1)$ as a constant point on the estimation line [15]. Typically, the I-frame is excluded from the quality control. At the beginning, we empirically initialize $a^{(1)}$ and $b^{(1)}$ as 2.16 and -1.16, respectively. $a^{(p+1)}$ and $b^{(p+1)}$ are updated as follows,

$$\begin{aligned} a^{(p+1)} &= \frac{1 - TM_{RR}^{(p)}}{1 - SSIM^{(p)}}, \\ b^{(p+1)} &= 1 - a^{(p+1)}. \end{aligned} \tag{11}$$

Here, $SSIM^{(p)}$ denotes the actual SSIM value of the p-th encoded frame, which is accessible when ecoding the $(p+1)$-th frame

III. Experimental Results

The proposed quality control method is validated based on the state-of-the-art VVC reference software VTM-9.0 [22]. Low delay P (LDP) and low delay B (LDB) configurations are considered in the experiment. The base QPs are set as 22, 27, 32, 37 following the common test conditions (CTC) [23]. In particular, we treat the averaged SSIM value $SSIM_t$ with the fixed-QP coding scheme as the target quality for each sequence. The comparisons regarding the SSIM variations at frame level with and without the proposed quality control scheme are shown in Fig. 2 and Fig. 3. Due to the fact that the perceptual quality measure SSIM is highly dependent on the video content, it can be observed that the quality at frame or GoP level highly deviates from the target quality for both higher bitrate (QP=22) and lower bitrate (QP=37) coding scenarios, owing to the variations of the video contents. With the guidance of the proposed quality control scheme, the quality variation becomes smoother, centralizing to the target SSIM level. Fig. 2(b), Fig. 2(d), Fig. 3(b) and Fig. 3(d) provide the absolute differences between the target SSIM and encoded SSIM for each frame where the proposed method achieves substantially lower quality deviations.

To further verify that the proposed quality control scheme could effectively regulate the encoding quality, quantification results regarding the SSIM variations with and without the proposed quality control scheme is measured, which can be formulated as follows,

$$\begin{aligned} V_{anc} &= \frac{\sum_{j=1}^{N_f} |SSIM_{anc}^{(j)} - SSIM_t|}{N_f}, \\ V_{pro} &= \frac{\sum_{j=1}^{N_f} |SSIM_{pro}^{(j)} - SSIM_t|}{N_f}, \end{aligned} \tag{12}$$

where N_f is the total frame number. $SSIM_{pro}^{(j)}$ and $SSIM_{anc}^{(j)}$ denotes the SSIM values of the j-th encoded frame with and without the proposed quality control scheme, respectively. The decrease of the SSIM variations when cooperating with the proposed scheme can be described as,

$$DV = \frac{V_{pro} - V_{anc}}{V_{anc}} \times 100\%. \tag{13}$$

The results of V_{anc}, V_{pro} and DV are illustrated in Table I. The proposed method could reduce over 60% of the SSIM variations under LDP and LDB configurations, illustrating the better match of the desired encoded quality level.

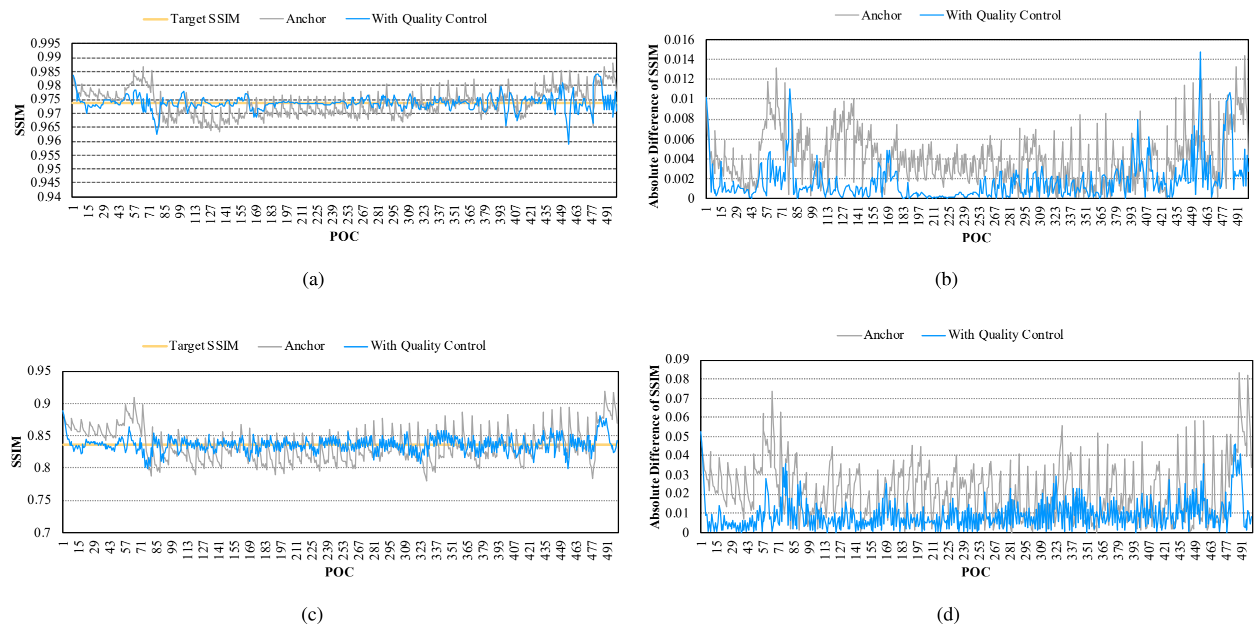

Fig. 2. SSIM values and absolute differences for each encoded frame in “BasketballPass” with and without the proposed quality control. (a) SSIM values, QP=22; (b) the absolute SSIM differences, QP=22; (c) SSIM values, QP=37; (d) the absolute SSIM differences, QP=37.

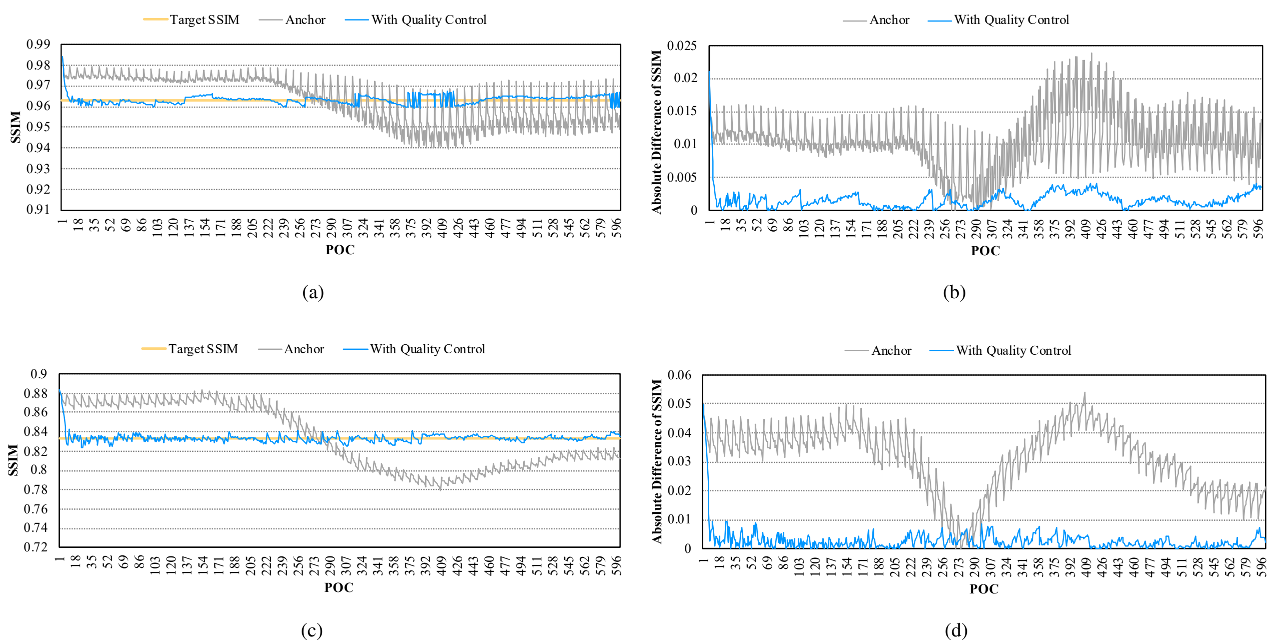

Fig. 3. SSIM values and absolute differences for each encoded frame in “BQSquare” with and without the proposed quality control. (a) SSIM values, QP=22; (b) the absolute SSIM differences, QP=22; (c) SSIM values, QP=37; (d) the absolute SSIM differences, QP=37.

IV. Conclusions

In this paper, we have systematically studied the quality control based on SSIM in VVC. The design philosophy is to impose the target quality constraint instead of the bitrate constraint on the encoded videos, in an effort to produce the video bitstream the quality of which could precisely reach the desired quality level. Moreover, the quality smoothness prior at the frame level could further benefit the quality control, leading to an accurate, consistent and readily plugged-in framework that could be widely deployed in many application

TABLE I
ILLUSTRATION OF THE V_{anc}, V_{pro} AND DV WITH AND WITHOUT THE PROPOSED QUALITY CONTROL SCHEME UNDER LDP AND LDB CONFIGURATIONS

Class	Sequence	Configuration	LDP				LDB			
		QP	22	27	32	37	22	27	32	37
C	BasketballDrill	V_{anc}	0.0019	0.0039	0.0055	0.0058	0.0019	0.0039	0.0053	0.0058
		V_{pro}	0.0008	0.0012	0.0021	0.0033	0.0008	0.0012	0.0021	0.0035
		DV	-58.6%	-69.4%	-60.9%	-43.1%	-59.0%	-69.6%	-60.0%	-39.0%
	BQMall	V_{anc}	0.0029	0.0051	0.0091	0.0139	0.0028	0.0050	0.0089	0.0138
		V_{pro}	0.0007	0.0010	0.0021	0.0037	0.0007	0.0010	0.0021	0.0033
		DV	-74.9%	-80.3%	-77.1%	-73.6%	-75.6%	-80.2%	-76.8%	-76.3%
	PartyScene	V_{anc}	0.0027	0.0048	0.0077	0.0104	0.0025	0.0046	0.0074	0.0102
		V_{pro}	0.0011	0.0016	0.0025	0.0030	0.0012	0.0017	0.0024	0.0030
		DV	-58.5%	-66.6%	-68.1%	-70.9%	-54.1%	-62.5%	-67.5%	-70.9%
	RaceHorses	V_{anc}	0.0052	0.0098	0.0199	0.0389	0.0049	0.0092	0.0193	0.0384
		V_{pro}	0.0013	0.0029	0.0061	0.0089	0.0013	0.0030	0.0057	0.0090
		DV	-75.2%	-70.3%	-69.5%	-77.1%	-74.6%	-67.1%	-70.3%	-76.5%
D	BasketballPass	V_{anc}	0.0040	0.0087	0.0151	0.0214	0.0040	0.0086	0.0149	0.0209
		V_{pro}	0.0018	0.0036	0.0063	0.0092	0.0017	0.0038	0.0061	0.0097
		DV	-56.6%	-59.2%	-57.9%	-57.0%	-57.7%	-56.3%	-59.1%	-53.8%
	BQSquare	V_{anc}	0.0104	0.0183	0.0253	0.0312	0.0106	0.0182	0.0247	0.0314
		V_{pro}	0.0016	0.0019	0.0022	0.0027	0.0014	0.0019	0.0022	0.0026
		DV	-84.8%	-89.6%	-91.3%	-91.2%	-86.6%	-89.8%	-91.1%	-91.7%
	BlowingBubbles	V_{anc}	0.0044	0.0075	0.0104	0.0128	0.0043	0.0073	0.0102	0.0125
		V_{pro}	0.0015	0.0020	0.0032	0.0041	0.0013	0.0018	0.0032	0.0038
		DV	-66.5%	-74.0%	-69.0%	-67.8%	-70.7%	-74.9%	-69.0%	-69.6%
	RaceHorses	V_{anc}	0.0039	0.0114	0.0271	0.0516	0.0036	0.0112	0.0268	0.0515
		V_{pro}	0.0013	0.0035	0.0072	0.0094	0.0012	0.0034	0.0072	0.0089
		DV	-67.0%	-69.6%	-73.5%	-81.8%	-67.1%	-69.4%	-73.1%	-82.6%
E	FourPeople	V_{anc}	0.0009	0.0013	0.0021	0.0030	0.0008	0.0012	0.0021	0.0030
		V_{pro}	0.0003	0.0004	0.0008	0.0013	0.0003	0.0004	0.0008	0.0012
		DV	-62.7%	-65.9%	-63.6%	-57.9%	-63.8%	-65.0%	-59.9%	-58.8%
	Johnny	V_{anc}	0.0012	0.0011	0.0010	0.0013	0.0012	0.0011	0.0010	0.0212
		V_{pro}	0.0004	0.0004	0.0006	0.0009	0.0003	0.0004	0.0006	0.0095
		DV	-70.0%	-63.7%	-39.9%	-30.0%	-74.8%	-65.9%	-41.3%	-55.3%
	KristenAndSara	V_{anc}	0.0010	0.0014	0.0018	0.0021	0.0009	0.0014	0.0018	0.0021
		V_{pro}	0.0007	0.0007	0.0009	0.0012	0.0007	0.0007	0.0009	0.0012
		DV	-28.0%	-49.7%	-49.1%	-43.8%	-27.5%	-49.8%	-51.4%	-42.4%
Averaged DV			-63.9%	-68.9%	-65.4%	-63.1%	-64.7%	-68.2%	-65.4%	-65.2%

scenarios. The proposed scheme is accomplished with a well established distortion-quantization model that is built upon a reduced-reference quality assessment model, and extensive experimental results show that the proposed quality control scheme reduces more than 60% of the quality variations for encoded videos.

ACKNOWLEDGMENT

This work was supported in part by Hong Kong ITF UICP under Grant 9440203, in part by the Hong Kong RGC Early Career Scheme under Grant 9048122 (CityU 21211018), and in part by the City University of Hong Kong Applied Research Grant under Grant 9667192.

REFERENCES

[1] S. Liu Y.-K. Wang B. Bross, J. Chen, "Versatile video coding (draft 9)," *JVET-R2001*, 2020.

[2] G. J. Sullivan, J. Ohm, W. Han, and T. Wiegand, "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] X. Li, H.-C. Chuang, J. Chen, M. Karczewicz, L. Zhang, X. Zhao, and A. Said, "Multi-type-tree," *Joint Video Exploration Team (JVET), doc. JVET-D0117*, 2016.

[4] B. Bross, H. Schwarz, D. Marpe, and T. Wiegand, "CE 3.3 related: Wide angular intra prediction for non-square blocks," *Joint Video Exploration Team (JVET), doc. JVET-L0279*, Oct. 2018.

[5] S. De-Luxán-Hernández, V. George, J. Ma, T. Nguyen, H. Schwarz, D. Marpe, and T. Wiegand, "An intra subpartition coding mode for VVC," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 1203–1207.

[6] L. Zhang, K. Zhang, H. Liu, H. C. Chuang, Y. Wang, J. Xu, P. Zhao, and D. Hong, "History-based motion vector prediction in versatile video coding," in *2019 Data Compression Conference (DCC)*, 2019, pp. 43–52.

[7] X. Zhao, J. Chen, M. Karczewicz, A. Said, and V. Seregin, "Joint separable and non-separable transforms for next-generation video coding," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2514–2525, 2018.

[8] H. Schwarz, T. Nguyen, D. Marpe, and T. Wiegand, "Hybrid video

coding with trellis-coded quantization," in *2019 Data Compression Conference (DCC)*, March 2019, pp. 182–191.

[9] M. Wien and B. Bross, "Tutorial 4: Versatile video coding – algorithms and specification," https://2020.ieeeicme-virtual.org/presentation/tutorial/tutorial-4-versatile-video-coding-algorithms-and-specification, 2020.

[10] X. Zhang, C. Yang, H. Wang, W. Xu, and C. . J. Kuo, "Satisfied-user-ratio modeling for compressed video," *IEEE Transactions on Image Processing*, vol. 29, pp. 3777–3789, 2020.

[11] Z. Wang and A. Rehman, "Begin with the end in mind: A unified end-to-end quality-of-experience monitoring, optimization, and management framework," *SMPTE Motion Imaging Journal*, vol. 128, no. 2, pp. 1–8, 2019.

[12] Z. Wang, K. Zeng, A. Rehman, H. Yeganeh, and S. Wang, "Objective video presentation QoE predictor for smart adaptive video streaming," in *Applications of Digital Image Processing XXXVIII*. International Society for Optics and Photonics, 2015, vol. 9599, p. 95990Y.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[14] S. Wang, A. Rehman, K. Zeng, J. Wang, and Z. Wang, "SSIM-motivated two-pass VBR coding for HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 10, pp. 2189–2203, 2017.

[15] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 4, pp. 516–529, 2012.

[16] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Perceptual video coding based on SSIM-inspired divisive normalization," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1418–1429, 2013.

[17] A. C. Brooks, X. Zhao, and T. N. Pappas, "Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions," *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1261–1273, 2008.

[18] A. Rehman and Z. Wang, "Reduced-reference ssim estimation," in *2010 IEEE International Conference on Image Processing*, 2010, pp. 289–292.

[19] S. S. Channappayya, A. C. Bovik, and R. W. Heath, "Rate bounds on ssim index of quantized images," *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1624–1639, 2008.

[20] X. Li, N. Oertel, A. Hutter, and A. Kaup, "Laplace distribution based lagrangian rate distortion optimization for hybrid video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 2, pp. 193–205, 2009.

[21] G. J. Sullivan, "Adaptive quantization encoding technique using an equal expected-value rule," *Joint Video Team of ISO/IEC and ITU-T, doc. JVT-N011*, Jan. 2005.

[22] "VVCSoftware_VTM repository," https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tree/VTM-9.0, 2020.

[23] F. Bossen, J. Boyce, X. Li, V. Seregin, and K. Suhring, "JVET common test conditions and software reference configurations for SDR video," *Joint Video Exploration Team (JVET), doc. JVET-N1010*, 2019.