On SSIM-Bit Rate Comparison of HEVC Encoders

Tiesong Zhao*, Zhou Wang†, Sam Kwong‡, and Chang Wen Chen§
* College of Physics and Information Engineering, Fuzhou University, Fuzhou, China
† Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada
‡ Department of Computer Science, City University of Hong Kong, Hong Kong
§ Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, New York
Emails: ztiesong@uwaterloo.ca, zhouwang@ieee.org, cssamk@cityu.edu.hk, chencw@buffalo.edu

Abstract—**The popular Structural SIMilarity (SSIM) index has shown to be a good perceptual criterion for testing and optimizing video encoders such as the MPEG-H/H.265 High Efficiency Video Coding (HEVC). However, it is still unclear how to compare two HEVC encoders with a number of bit rates and SSIM values. In this work, we study the video quality comparison of HEVC encoders based on the Bit Rate-SSIM (R-S) curves. Our contributions are three-fold. First, we exploit how to interpolate an R-S curve when only a small number of R-S samples are available. We also develop to minimize the weighted sum of squared errors during the interpolation process, where the weights are pre-defined by users to satisfy the diversified requirements. Second, we propose an approach to evaluate the perceptual coding gain in terms of the R-S improvement between two individual video encoders. Third, as an application, we utilize the proposed approach to compare the R-S performances of HEVC encoders with different configurations.**

I. Introduction

During the recent years, the increasing requirements of video services have greatly promoted the development of video coding standards, among which the most recent milestone is the new MPEG-H/H.265 High Efficient Video Coding (HEVC) technique [1]. Essentially in all popular video coding standards including HEVC and H.264 [2], the lossy video coding problem is formulated as a Rate-Distortion Optimization (RDO) problem to minimize the coding distortion subject to a constraint on Bit Rate (BR). Nevertheless, the current RDO approach has been criticized for its definition of distortion. The traditional objective distortion/quality measures such as Sum of Absolute Difference (SAD), Mean Squared Error (MSE), and Peak-Signal-to-Noise-Ratio (PSNR), have been found to be poorly correlated to the perceived video quality of Human Vision System (HVS).

To address this issue, perceptual video coding techniques have been developed in which subjective distortion/quality measures are used to imitate the human perception of video quality. As a criterion of these subjective measures, Mean Opinion Score (MOS) or its variation like differential MOS (DMOS) is obtained though subjective test and has been believed to reflect the human perceived quality [3], [4], [5]. However, it is still impossible to integrate the subjective test into real-life video encoders. As an alternate, the perceptual Video Quality Assessment (VQA) technique has undergone significant development aiming at designing a better visual quality metric. Generally, the VQA approaches can be categorized into two groups, namely, vision modeling approaches and engineering approaches, among which the bottom-up engineering approaches have been becoming more popular [6].

In engineering approaches, the picture contents are analyzed to extract features and artifacts for the evaluation of video distortion. The most popular engineering approaches include Video Quality Metric (VQM) [7], the Structural SIMilarity (SSIM) index family [8], [9], [10], [11], and the MOtion-based Video Integrity Evaluation (MOVIE) index [12]. Among them, SSIM has been widely employed as a good successor of traditionally used MSE/PSNR in H.264 mode selection [13], [14], Motion Estimation (ME) [15], RDO [16], [17], rate control [18], [19], quantization [14], [20], and so on. Recently, the divisive normalization method used in [14] has also been exploited in HEVC to design mode selection and RDO algorithms [21]. SSIM has also been employed as a quality measure in other perceptual coding methods like [22], where a perceptual RDO method was designed for HEVC.

To show the perceptual improvement in terms of SSIM, the intuitive BR-SSIM (R-S) curves were presented in [14], [16], [21]. Aiming at numerically estimating the average R-S performance gain over a large range of BRs, Bjontegaard's average PSNR increase (BDPSNR) and BR increase (BDBR) [23] have been used in [14], [17], [18], [21], [22]. By using the PSNR-based measure to evaluate the R-S performance, they employed an implicit assumption that the R-S curves had similar characteristics to the traditionally used RD curves (*i.e.*, PSNR versus BR curves). However, the R-S curves have different shapes to the RD curves [24], and therefore the aforementioned BDPSNR & BDBR approach may not be an ideal scheme to calculate the perceptual improvement.

In this paper, we attempt to address the main issues regarding the comparison of SSIM, which is a recommended quality criterion to be embedded into the video encoder [24]. First, by comparing the differences between the R-S and RD curves, we discuss why we need to develop a new comparison approach for the R-S curves. Then, we propose to interpolate R-S curves and to compare the coding performances of two SSIM-inspired encoders. We also develop a scheme to interpolate and compare the R-S curves when the user-defined weights are employed to minimize the weighted sum of square errors in R-S curve interpolation. Finally, we utilize the proposed approach to compare the R-S performances of HEVC encoders.

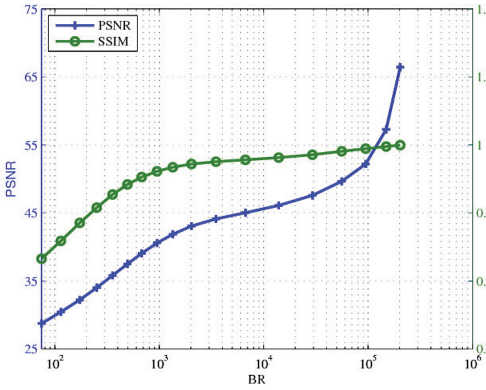


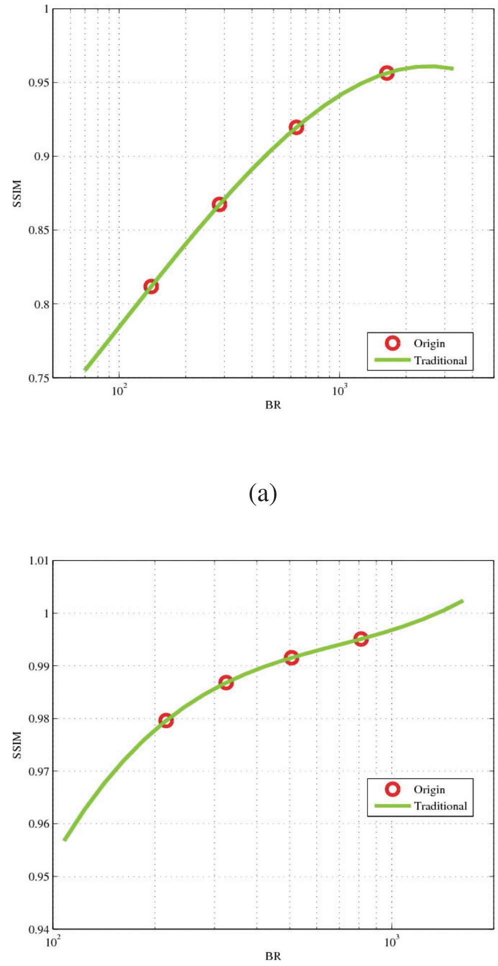
Fig. 1. A typical example of PSNR and SSIM versus BR [24].

II. Why Interpolate R-S Curves?

A VQA measure is utilized in video coding as both a *testing tool* and an *optimization tool*. As a *testing tool*, the VQA measure is used to choose the best video from a group of coded videos, or determine the best video coding configuration or algorithm from multiple choices. As an *optimization tool*, the VQA measure is embedded into the design and optimization of video codecs to improve the compression performance. To fulfill these tasks, there are several attributes required for this measure, which include *correlated with perceptual quality*, *low complexity*, *good mathematical properties*, *localization in quality prediction*, *saturated at high rate* and its *popularity* [24]. When examined by these desirable attributes, SSIM deserves a score of "Good" or above in all attributes, which makes SSIM and its derivatives or direct improvements upon them (*e.g.*, by incorporating temporal assessment approaches) be the top choice in the state-of-the-art perceptual video coding.

Regarding to the attribute *saturated at high rate* as a *testing tool*, SSIM index behaves very different from PSNR and this fact leads to different curves as shown in Fig. 1. The RD curve measured by PSNR and BR behaves approximately a convex function and does not saturate at high BR; in contrast, the R-S curve can be illustrated by a concave function which saturates to 1 at high BR end. Therefore, the interpolation method of RD curve may not be applicable to R-S curves.

Until now, the most popular RD curve interpolation method was developed in [23], where the RD curve was interpolated by polynomial functions. PSNR was approximated by a cubic polynomial interpolation of logarithmic BR. Follow this, we may interpolate R-S curves by simply replacing PSNR by SSIM, which has been utilized in some recent works. However, this interpolation does not guarantee a saturated SSIM at high BR end. With this method SSIM may divergent to infinity, which is against to both the characteristics of SSIM and the convergence rule to be a good VQA measure. To examine this, we interpolate R-S points of the HEVC Common Test Condition (CTC) sequences [25] and observe the corresponding performances at low and high BR ends. The Random Access (RA) configuration and four Quantization parameters (Qps), including 22, 27, 32 and 37, are used. Non-overlapping blocks [24] are adopted to calculate the SSIM values with a window of 4×4.



(a)

(b)

Fig. 2. SSIM interpolation errors when using the traditional method. (a) *BQSquare* (416×240). (b) *SlideShow* (1280×720).

According to the observation, the traditional interpolation method cannot work well for some CTC the sequences, especially at the high BR end. Two examples are presented in Fig. 2. For the Class D sequence *BQSquare* (416×240), SSIM is predicted to decrease when BR is larger than some threshold. In other words, to allocate more bits may result in a lower video quality in terms of SSIM, which is untrue in video coding. For the Class F sequence *SlideShow* (1280×720), it is predicted that SSIM can be larger than 1, which exceeds the range of an SSIM value. Therefore, it is necessary to develop an improved interpolation method to avoid these problems.

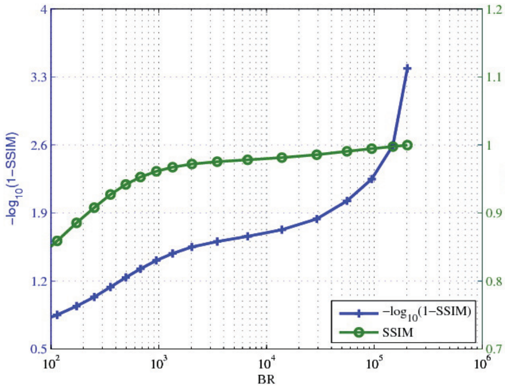


Fig. 3. SSIM transform to logarithmic domain.

III. The Interpolation and Comparison of R-S Curves

To utilize the above PSNR-based method to interpolate an R-S curve in terms of SSIM and BR, a straightforward method is to transform the R-S curve to a similar form of the RD curve shown in Fig. 1. Here we utilize logarithmic transforms of SSIM and BR, as

$$\mathrm{S}_{\log} = -\log_{10}(1-S), \tag{1}$$

$$\mathrm{R}_{\log} = \log_{10} R, \tag{2}$$

where S and R denote SSIM and BR, respectively.

An example of $\mathrm{S}_{\log}$ versus $\mathrm{R}_{\log}$ is given in Fig. 3, where source data is the same to that in Fig. 1. It is noticed that $\mathrm{S}_{\log}$ and PSNR have very similar shapes as functions of $\mathrm{R}_{\log}$. Thus we can use a cubic polynomial interpolation as

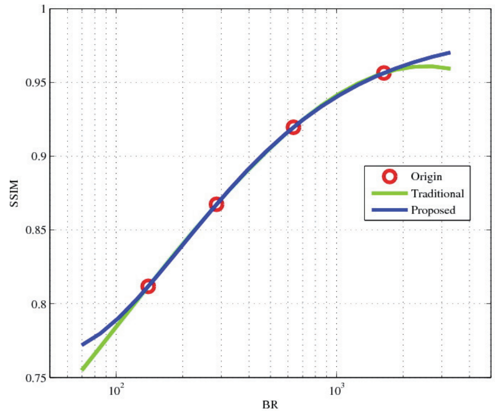
$$\mathrm{S}_{\log} = \sum_{n=1}^{N} \alpha_n \cdot {\mathrm{R}_{\log}}^{N-n}, \tag{3}$$

where N and $\alpha_n, n = 1, 2, ...N$ are the order and coefficients of polynomial functions, respectively. $\alpha_n, n = 1, 2, ...N$ can be obtained by minimizing the mean square error of all S_{log} samples. Generally $\mathrm{S}_{\log} > 0$. Similar to [23], we set $N = 3$ for a cubic polynomial function. Therefore, an R-S curve can be interpolated with at least four R-S points.

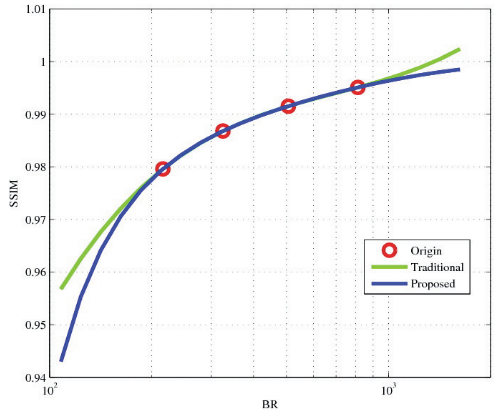
Substituting Eqs. (1) and (2) into Eq. (3), we obtain the interpolated SSIM as

$$S = 1 - 10^{-\sum_{n=1}^{N} \alpha_n \cdot (\log_{10} \mathrm{R})^{N-n}}, \tag{4}$$

where $S \in (0, 1)$ for lossy video encoders. With the proposed interpolation method, the R-S prediction errors shown in Fig. 2 can be prevented. It can be seen in Fig. 4 that for both the above two CTC sequences, the propose prediction method converges to 1 at high BR end.



(a)



(b)

Fig. 4. SSIM prediction errors can be avoided with the proposed method. (a) *BQSquare* (416×240). (b) *SlideShow* (1280×720).

On the other hand, $\mathrm{R}_{\log}$ can also be interpolated as a polynomial function of $\mathrm{S}_{\log}$ and thus

$$R = 10^{\sum_{n=1}^{N} \beta_n \cdot (-\log_{10}(1-S))^{N-n}}, \tag{5}$$

where $\beta_n, n = 1, 2, ...N$ are the interpolation coefficients with least square errors of all samples of R. We also set $N = 3$ for a cubic interpolation.

Besides the aforementioned improvement at high BR end, the proposed interpolation can also reduce the Root Mean Squared Error (RMSE) when fitting the curves. To validate this, we encode all CTC sequences with HEVC reference software HM 16.0 and the configuration of JCTVC-L1100 [25]. Four configurations are employed including intra (IA), RA, low delay (LD) and low delay P (LP). Four Qps are coded including 22, 27, 32 and 37. The Motion Estimation (ME) search ranges for 416×240, 832×480, 1024×768, 1280×960,

TABLE I
COMPARISON OF TRADITIONAL AND PROPOSED METHODS IN TERMS OF PREDICTION ERRORS.

Video Sequences		IA		RA		LD		LP	
		RMSE_S	RMSE_R	RMSE_S	RMSE_R	RMSE_S	RMSE_R	RMSE_S	RMSE_R
Class A	Traditional	5.86e-15	1.36e-07	4.41e-15	2.58e-08	-	-	-	-
	Proposed	6.75e-15	8.70e-10	4.03e-15	7.40e-10	-	-	-	-
Class B	Traditional	2.83e-15	2.38e-08	5.94e-16	3.55e-08	4.94e-16	1.06e-08	4.93e-16	2.15e-08
	Proposed	4.43e-15	2.03e-10	4.04e-16	9.70e-10	2.25e-16	3.90e-10	3.62e-16	2.42e-10
Class C	Traditional	1.56e-15	4.27e-09	1.28e-15	1.73e-09	6.10e-16	8.33e-10	7.95e-16	9.42e-10
	Proposed	2.84e-15	5.33e-11	2.48e-16	7.78e-12	2.05e-16	6.79e-12	1.75e-16	6.31e-12
Class D	Traditional	2.07e-15	6.60e-10	2.42e-16	1.48e-10	2.39e-16	1.09e-10	2.96e-16	9.87e-11
	Proposed	2.96e-15	1.43e-11	2.30e-16	2.19e-12	2.24e-16	1.36e-12	3.73e-16	5.45e-12
Class E	Traditional	2.69e-15	5.16e-08	-	-	5.78e-16	2.36e-08	6.97e-16	2.11e-08
	Proposed	1.37e-16	1.66e-11	-	-	6.90e-17	4.89e-11	1.64e-16	1.79e-11
Class F	Traditional	5.05e-15	2.37e-07	9.22e-16	4.51e-08	9.62e-16	2.04e-08	9.82e-16	1.00e-08
	Proposed	2.44e-15	3.73e-11	3.80e-16	3.63e-12	7.58e-16	1.14e-11	3.67e-16	3.50e-12
Average	**Traditional**	**3.35e-15**	**7.44e-08**	**1.45e-15**	**2.23e-08**	**5.72e-16**	**1.05e-08**	**6.42e-16**	**1.07e-08**
	Proposed	**3.44e-15**	**2.07e-10**	**1.03e-15**	**3.75e-10**	**3.04e-16**	**1.09e-10**	**2.98e-16**	**6.62e-11**

1920×1080 and 2560×1600 sequences are set to be 16, 32, 64, 64, 128 and 256, respectively. The average RMSE values of R and S for all CTC classes as well as the average values of all CTC sequences are presented in Table I (LD/LP for Class A and RA for Class E are not specified in [25] and thus they are not given). From the table, the proposed interpolation method can further reduce the prediction errors of R-S points when comparing with the traditional method, especially for the prediction of BR. This fact ulteriorly demonstrates the prediction accuracy of the proposed interpolation method.

Based on the interpolated R-S curves, we can compare the R-S performances between two encoders. We propose two measures including Average Delta SSIM (ADSSIM) for average SSIM increase with the same BR, and Average Delta BR (ADBR) for BR increase whilst keeping the same SSIM performance. ADSSIM can be obtained as follows. First, we denote Eq. (4) as

$$\mathrm{S} = f(\mathrm{R}_{\log}); \tag{6}$$

Then, we interpolate two R-S curves $f_1(\cdot)$ and $f_2(\cdot)$ for two encoders; Finally, with a user-defined logarithmic BR range $[\mathrm{R}_{log,min}, \mathrm{R}_{log,max}]$, we can get ADSSIM as

$$\begin{aligned}\mathrm{ADSSIM} =& \frac{1}{\mathrm{R}_{log,max} - \mathrm{R}_{log,min}} \cdot \\ & \int_{\mathrm{R}_{log,min}}^{\mathrm{R}_{log,max}} \{f_1(R_{log}) - f_2(R_{log})\} dR_{log}.\end{aligned} \tag{7}$$

ADBR can be obtained in a similar way. By transforming Eq. (5) into

$$\mathrm{R}_{\log} = g(\mathrm{S}_{\log}), \tag{8}$$

ADBR can be derived as

$$\begin{aligned}\mathrm{ADBR}_{\log} =& \frac{1}{\mathrm{S}_{log,max} - \mathrm{S}_{log,min}} \cdot \\ & \int_{\mathrm{S}_{log,min}}^{\mathrm{S}_{log,max}} \{g_1(S_{log}) - g_2(S_{log})\} dS_{log},\end{aligned} \tag{9}$$

$$\mathrm{ADBR} = (10^{ADBR_{log}} - 1) \times 100\%, \tag{10}$$

where $\mathrm{S}_{log,min}$ and $\mathrm{S}_{log,max}$ are the user-defined minimum and maximum S values transformed by Eq. (1). The transform of Eq.(10) aims at obtaining an average BR increase in terms of percentage, which is of the same form to BDBR.

In Eqs. (4) and (5) the interpolation coefficients $\alpha_n, n = 1, 2, ..., N$ and $\beta_n, n = 1, 2, ..., N$ are obtained by minimizing the mean squared errors during the curve fitting. However, different R-S points may not be equally treated in some applications. In some situations, the users care more about low BR coding performance; while in some other cases, the high BR end is more important. From this point, we derive the above coefficients by minimizing the sum of weighted least squares. Assume that there are $N \geq 4$ R-S points available, as (S_1, R_1), (S_2, R_2), ... , (S_N, R_N), and let the weights be $w_{S,1}$, $w_{S,2}$, ..., $w_{S,N}$, the objective to fit the R-S curve is

$$\min\left\{\sum_{k=1}^{M} w_{S,k} \cdot (\mathrm{S}_k - \mathrm{S}_k^*)^2\right\}, \tag{11}$$

where S_k^* is the predicted value of S_k with R-S curve fitting. Correspondingly, in logarithmic domain, the objective is

$$\min\left\{\sum_{k=1}^{M} w'_{S,k} \cdot (\mathrm{S}_{\mathrm{log,k}} - \mathrm{S}^*_{\mathrm{log,k}})^2\right\}, \tag{12}$$

where $\mathrm{S}_{\mathrm{log,k}}$ and $\mathrm{S}^*_{\mathrm{log,k}}$ are S_k and S^*_k transformed by Eq. (1), respectively. $w'_{S,k}$ is the corresponding weight of $w_{S,k}$ in logarithmic domain, and

$$w_{S,k} \cdot (\partial S)^2|_{S=S_k} = w'_{S,k} \cdot (\partial \mathrm{S}_{\log})^2|_{\mathrm{S}_{\log}=\mathrm{S}_{\log,k}}. \tag{13}$$

By substituting Eq. (1) and excluding constants,

$$w'_{S,k} = w_{S,k} \cdot (1 - S_k)^2. \tag{14}$$

Similarly, we get

$$w'_{R,k} = w_{R,k} \cdot R_k^2. \tag{15}$$

With the user-defined $w_{S,k}, k = 1, 2, ..., N$ and $w_{R,k}, k = 1, 2, ..., N$, we can transform them into $w'_{S,k}, k = 1, 2, ..., N$ and $w'_{R,k}, k = 1, 2, ..., N$ by using Eqs. (14) and (15). The new

TABLE II
COMPARISON OF DIFFERENT CONFIGURATIONS OF HEVC.

Video sequences		LD vs. LP		RA vs. LP		RA vs. LD	
		ADSSIM	ADBR	ADSSIM	ADBR	ADSSIM	ADBR
Class B	*Kimono* 1920×1080	0.0039	-11.70	0.0085	-21.11	0.0046	-10.09
	ParkScene 1920×1080	0.0031	-7.95	0.0067	-13.58	0.0038	-6.57
	Cactus 1920×1080	0.0032	-9.64	0.0043	2.19	0.0012	13.25
	BQTerrace 1920×1080	0.0064	-13.32	0.0054	2.08	-0.0009	19.13
	BasketballDrive 1920×1080	0.0038	-10.46	0.0068	-1.45	0.0031	9.88
	Average	0.0041	-10.61	0.0064	-6.37	0.0024	5.12
Class C	*RaceHorses* 832×480	0.0036	-6.87	0.0064	-7.94	0.0028	-0.89
	BQMall 832×480	0.0033	-8.65	0.0131	-21.56	0.0098	-14.43
	PartyScene 832×480	0.0036	-10.23	0.0196	-24.05	0.0160	-15.48
	BasketballDrill 832×480	0.0042	-9.16	0.0102	-19.70	0.0061	-11.95
	Average	0.0037	-8.73	0.0123	-18.31	0.0087	-10.69
Class D	*RaceHorses* 416×240	0.0020	-4.35	0.0080	-10.29	0.0060	-6.43
	BQSquare 416×240	0.0053	-13.50	0.0169	-32.20	0.0116	-21.31
	BlowingBubbles 416×240	0.0017	-7.72	0.0201	-22.94	0.0184	-16.74
	BasketballPass 416×240	0.0031	-4.59	0.0138	-18.68	0.0106	-14.76
	Average	0.0030	-7.54	0.0147	-21.03	0.0117	-14.81
Class E	*FourPeople* 1280×720	0.0015	-8.49	-	-	-	-
	Johnny 1280×720	0.0018	-15.02	-	-	-	-
	KristenAndSara 1280×720	0.0014	-9.62	-	-	-	-
	Average	0.0016	-11.04	-	-	-	-
Class F	*BaskeballDrillText* 832×480	0.0044	-9.94	0.0107	-20.68	0.0062	-12.18
	ChinaSpeed 1024×768	0.0017	-3.00	-0.0011	1.91	-0.0026	4.96
	SlideEditing 1280×720	0.0006	-0.39	-0.0745	294.87	-0.0751	296.12
	SlideShow 1280×720	0.0005	-2.47	-0.0015	-0.35	-0.0020	2.56
	Average	0.0018	-3.95	-0.0166	68.94	-0.0184	72.86
Average		**0.0030**	**-8.35**	**0.0043**	**5.09**	**0.0012**	**12.65**

weights are then utilized to obtain the weighted sum of prediction errors, by minimizing which we could get the coefficients $\alpha_n, n = 1, 2, ..., N$ and $\beta_n, n = 1, 2, ..., N$ to obtain ADSSIM and ADBR. Typically, we can use $w_{S,k} = 1$ for ADSSIM (*i.e.* identical weight for all SSIM values) and $w_{R,k} = R_k^{-2}$ (*i.e.* $w'_{R,k} = 1$, identical weight for all logarithmic BR values) for ADBR. These weights can be adjusted by users to fulfill diversified real-life requirements.

IV. THE APPLICATIONS OF ADSSIM AND ADBR

As an application, we compare the R-S performances of HEVC encoder HM 16.0 with different configurations: LD, LP and RA. To ensure overlapped R-S curves, eight R-S points are evaluated with Qps (12, 17, 22, 27, 32, 37, 42 and 47). The search range and SSIM calculation are set to be the same to Table I. The coding parameters including frame rate, frames to be encoded, intra period and input bit depth are according to HEVC CTC [25] and other coding parameters are set as the defaults of HM encoder. The calculation range $[\mathrm{S}_{log,min}, \mathrm{S}_{log,max}]$ and $[\mathrm{R}_{log,min}, \mathrm{R}_{log,max}]$ are set as the overlapped regions of the corresponding R-S curves.

We summarize the obtained ADSSIM and ADBR values in Table II. Because the LD/LP configurations are not specified in Class A and RA configuration is not specified in Class E, all the comparisons of Class A and the RA comparisons of Class E are not presented. From the table, we could get several important conclusions for HEVC configurations. First, LD outperforms LP in all CTC sequences with resolutions from 416×240 to 1920×1080, which demonstrates the efficiency of B frames over P frames. The improvements for Class F are not so significant, which may be due to lower probabilities of bi-predictions in these sequences. Second, the newly adopted RA configuration outperforms LD/LP for most of the Classes and sequences, yet it may also result in a large ADBR increase. The reasons may be two-fold. On one hand, RA uses a larger Group-Of-Picture (GOP) size than LD/LP (*e.g.* the default GOP sizes of RA and LD/LP are 8 and 4, respectively) and a hierarchical bi-prediction structure, which lead to an improved R-S performance; on the other hand, RA adopts more I frames than LD/LP to ensure a randomly accessed video stream, which also produces more coding bits due to the inefficiency of I frame. The contradictory aspects could either improve or impair the R-S efficiency, depending on the video characteristics and RA period during the encoding process. An example is the sequence *SlideEditing*, in which there exists a large region of static background. Therefore, only a few number of I frames are adequate; RA shows a lower coding efficiency because it has more I frames. However, such a case is not common in real-life video sequences. By excluding this sequence, RA reduces 13.02% and 5.07% bit rates of LP and LD, respectively on average. In conclusion, it is still desirable to use RA in some applications to support a more efficient encoder.

Another observation on ADSSIM and ADBR is that, the absolute value of the former is usually much smaller than the latter for most of the coding sequences. On average, the LD configuration reduces 8.35% ADBR of LP, but results in only a 0.0030 ADSSIM increase, which is due to the convergence of SSIM values. This fact implies that with the same BR,

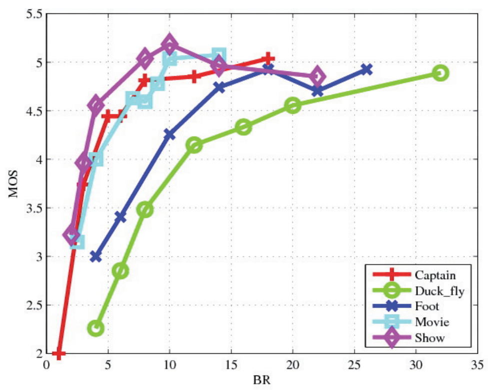


Fig. 5. Examples of MOS versus BR.

a small SSIM improvement may be still worthwhile because the corresponding ADBR is significant. Based on this, and also considering that the interpolation errors of BR is much smaller than those of SSIM (as shown in Table I), the ADBR, which indicates the average BR reduction with the same SSIM, may be more applicable in real-life encoders.

Besides the applications in comparing R-S curves, the proposed method may also be employed to calculate MOS (or MOS-like measure) versus BR curves because MOS have a high correlation to SSIM [26] and it also saturates at high BR end. Examples of MOS-BR curves are presented in Fig. 5, where the data is from IRCCyN/IVC 1080i database [27]. It is noted that the MOS values converges when the BR is larger than a threshold. Therefore, when a normalized MOS (*i.e.* the MOS value divided by the maximum MOS, which is between 0 and 1) is employed, the MOS-BR curve may also be interpolated with the proposed method and thus the MOS-BR performances of different encoders can be compared.

V. Conclusions

In this paper, we have discussed the drawbacks of the popular RD interpolation method when it is used to interpolate the R-S curves. We have developed a method to interpolate R-S curves and predict R-S performance of an encoder. Based on user-defined weights, we calculate ADSSIM and ADBR as measures of perceptual improvement of a video encoder when SSIM is utilized as the quality measure. The proposed ADSSIM and ADBR have wide applications in SSIM-based video coding optimization or other perceptual video coding schemes.

References

[1] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm, and T. Wiegand, "High efficiency video coding (HEVC) text specification draft 8, " Doc. JCTVC-J1003, July 2012.

[2] "Advanced video coding for generic audiovisual services," ISO/IEC 14496-10:2005(E) ITU-T Rec. H.264 (E), Mar. 2005.

[3] ITU-R Rec. BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," Jan. 2012.

[4] ITU-T Rec. P.910, "Subjective video quality assessment methods for multimedia applications," Apr. 2008.

[5] ITU-R Rec. BT.1788, "Methodology for the subjective assessment of video quality in multimedia applications," Jan. 2007.

[6] S. Winkler and P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 660-668, Sept. 2008.

[7] M.H. Pinson, and S. Wolf, "A new standardized method for objectively meausuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312-322, Sept. 2004.

[8] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81-84, Mar. 2002.

[9] Z. Wang, L. Lu, and A.C. Bovik, "Video Quality Assessment Based on Structural Distortion Measurement," *Signal Process.: Image Commun. special issue on Objective Video Quality Metrics*, vol. 19, no. 2, pp. 121-132, Feb. 2004.

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

[11] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *IEEE Asilomar Conf. Signals, Syst. and Comput.*, pp. 1398-1402, Nov. 2003.

[12] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335-350, Feb. 2010.

[13] Z.-Y. Mai, C.-L. Yang, and S.-L. Xie, "Improved best prediction mode(s) selection methods based on structural similarity in H.264 I-frame encoder," in *IEEE Int. Conf. Syst. Man and Cybern. 2005 (SMC'05)*, pp. 2673-2678, Oct. 2005.

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

[15] Z.-Y. Mai, C.-L. Yang, K.-Z. Kuang, and L.-M. Po, "A novel motion estimation method based on structural similarity for H.264 inter prediction," in *IEEE Int. Conf. Acoust. Speech and Singal Process. 2006 (ICASSP'06)*, pp. 913-916, May 2006.

[16] Y.-H. Huang, T.-S. Ou, P.-Y. Su and H.H. Chen, "Perceptual rate-distortion optimization using structural similarity index as quality metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1614-1624, Nov. 2010.

[17] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516-529, Apr. 2012.

[18] T.-S. Ou, Y.-H. Huang and H.H. Chen, "SSIM-Based Perceptual Rate Control for Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 5, pp. 682-691, May 2011.

[19] G.-L. Wu, Y.-J. Fu, S.-C. Huang and S.-Y. Chien, "Perceptual Quality-Regulable Video Coding System With Region-Based Rate Control Scheme," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2247-2258, Jun. 2013.

[20] Z. Chen and Y.-P. Tan, "Frame-level quantization control for perceptual quality constrained H.264/AVC video coding," in *IEEE Int. Symp. Circuits and Syst. 2011 (ISCAS'11)*, pp. 1231-1234, May 2011.

[21] A. Rehman and Z. Wang,"SSIM-Inspired Perceptual Video Coding for HEVC," *IEEE Int. Conf. Multimedia Expo 2012 (ICME'12)*, pp. 497-502, Jul. 2012.

[22] H. Zeng, K.N. Ngan and M. Wang, "Perceptual adaptive Lagrangian multiplier for high efficiency video coding," *Picture Coding Symp. 2013 (PCS'13)*, pp. 69-72, Dec. 2013.

[23] G. Bjontegaard, "Calculation of average PSNR difference between RD curves," *Proc. ITU-T Q.6/SG16 VCEG 13th Meeting*, Austin, TX, 2001.

[24] T. Zhao, K. Zeng, A. Rehman and Z. Wang, "On the Use of SSIM in HEVC," *IEEE Asilomar Conference on Signals, Systems and Computers 2013*.

[25] F. Bossen, "Common test conditions and software reference configurations,", Doc. JCTVC-L1100, ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Geneva, Jan, 2013.

[26] K. Zeng, A. Rehman, J. Wang and Z. Wang, "From H.264 to HEVC: Coding Gain Predicted by Objective Video Quality Assessment Models," VPQM 2013, Scottsdale, AZ, USA, Jan.-Feb. 2013.

[27] S. Pchard, R. Ppion, P. Le Callet, "Suitable methodology in subjective video quality assessment: a resolution dependent paradigm," *International Workshop on Image Media Quality and its Applications*, Kyoto, 2008.