IEEE International Workshop on Multimedia Signal Processing, Xiamen, China, Oct. 2015.

SSIM-Inspired Two-Pass Rate Control for High Efficiency Video Coding

Shiqi Wang, Abdul Rehman, Kai Zeng and Zhou Wang

Dept. of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON, Canada

***Abstract*—We propose a perceptual two-pass rate control scheme for High Efficiency Video Coding (HEVC). The target bits are optimally allocated by hierarchically constructing a perceptual uniform space derived based on an SSIM-inspired divisive normalization mechanism for each group of pictures (GoP), each frame, and each coding unit (CU). The Lagrange multiplier λ, which controls the trade-off between perceptual distortion and bit rate, is adopted as the GoP level complexity measure. After the first pass compression, Laplacian based rate and perceptual distortion models are established to adaptively derive λ, and the target bits are dynamically allocated by maintaining an uniform Lagrange multiplier level through λ equalization. Within each GoP, rate control is further performed at frame and CU levels in the perceptually uniform space. Extensive simulations verify that, the proposed scheme can achieve high accuracy rate control and superior rate-SSIM performance.**

***Index Terms*—Two-pass rate control, divisive normalization, SSIM index, High Efficiency Video Coding**

I. Introduction

Recently, there has been an exponentially increasing demand of high definition (HD) and beyond-HD videos, which has been creating an ever stronger demand for high performance video coding technologies. The high efficiency video coding (HEVC) standard [1], which was jointly developed by ITU-T Video Coding Expert Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG), was claimed to bring more than 50% coding gain compared to H.264/AVC. In HEVC, many novel coding techniques are developed. At the block level, an adaptive quadtree structure based on the coding tree unit (CTU) was employed, and three new concepts, named coding unit (CU), prediction unit (PU) and transform unit (TU), are introduced to specify the basic processing unit of coding, prediction and transform [2]. At the frame level, flexible reference management scheme based on the concept of reference frame set (RFS) was adopted to support the hierarchical coding structure [3].

Practically, to apply the video coding standards in real application scenarios, rate control schemes need to be incorporated into the encoder (for example, TM5 for MPEG-2, TMN8 for H.263 and VM8 for MPEG-4). In HEVC, several rate control algorithms are proposed, targeting at constant bit rate (CBR) coding. In [4], an adaptive rate control scheme was proposed by modeling the Rate-Quantization relationship with frame complexity, and Laplacian distribution based CTU level bit allocation is further developed to improve the coding performance. In [5], Lagrange parameter (λ) domain rate control was proposed, in which the quantization parameter (QP) value for each frame is obtained by the corresponding λ value. In [6], [7], considering the new reference frame selection mechanism, Rate-GOP based distortion and rate models were established and ρ domain rate control was proposed for HEVC, where ρ represents the percentage of zero coefficients in a frame after quantization.

Though these rate control algorithms have achieved substantial performance in control accuracy and coding performance, variable bit rate (VBR) coding of HEVC has not been fully investigated in the literature. In H.264/AVC, two-pass rate control for VBR coding have been intensively studied [8]–[10]. A general idea is to model scene complexity according to the first-pass statistics, and the quantization parameters for each frame can be derived according to the distributed bits. The task of optimal bit allocation can be converted into a typical Rate Distortion Optimization (RDO) problem to optimize the quality of the whole sequence. Central to such an optimization problem is the way in which the distortion D is defined, so that the overall quality of the whole video can approach what it is optimized for. This motivated us to employ the structural similarity (SSIM) index [11] as the image quality measure, which has been widely applied in various image/video processing areas due to its good compromise between quality evaluation accuracy and computation efficiency. In [12]–[14], SSIM-based RDO schemes were proposed, which shows good perceptual rate distortion performance. In [15], it is shown that substantial difference between SSIM and MSE lies in a locally adaptive divisive normalization process, which motivated us to develop divisive normalization based video coding schemes [16], [17] on the platform of H.264/AVC and HEVC.

In this paper, we propose a perceptual two-pass VBR scheme within the SSIM-inspired divisive normalization video coding framework. The RD performance is optimized by dynamically balancing the λ value for each frame, which is adaptively derived by establishing perceptual distortion and rate models. Constructively, adaptive GoP level, frame level and CTU level rate control schemes are proposed by transforming the prediction residuals into a perceptually uniform space.

II. SSIM-Based Divisive Normalization for Perceptual Video Coding

The proposed rate control scheme follows the divisive normalization based perceptual video coding approach [16],

978-1-4673-7478-1/15/$31.00 ©2015 IEEE

[17], in which the DCT transform coefficient of a residual block C_k is normalized with a positive normalization factor f:

$$C(k)' = C(k)/f. \tag{1}$$

As such, the quantization process of the normalized residuals for a given predefined Q_s can be formulated as

$$\begin{aligned} Q(k) &= \text{sign}\{C(k)'\}\text{round}\{\frac{|C(k)'|}{Q_s} + p\} \\ &= \text{sign}\{C(k)\}\text{round}\{\frac{|C(k)|}{Q_s \cdot f} + p\}, \end{aligned} \tag{2}$$

where p is the rounding offset in the quantization.

This implies that the quantization parameters for each coding unit can be adaptively adjusted according to the divisive normalization process. The factor f, which accounts for the perceptual importance, is derived from the SSIM index in DCT domain,

$$f_{dc} = \frac{\frac{1}{l}\sum_{i=1}^{l}\sqrt{X_i(0)^2 + Y_i(0)^2 + N \cdot C_1}}{E(\sqrt{X(0)^2 + Y(0)^2 + N \cdot C_1})} \tag{3}$$

$$f_{ac} = \frac{\frac{1}{l}\sum_{i=1}^{l}\sqrt{\frac{\sum_{k=1}^{N-1}(X_i(k)^2+Y_i(k)^2)}{N-1} + C_2}}{E(\sqrt{\frac{\sum_{k=1}^{N-1}(X(k)^2+Y(k)^2)}{N-1}} + C_2)}, \tag{4}$$

where X and Y represents the DCT coefficients for the original and distorted blocks. $E(\cdot)$ denotes the expectation quantity over the whole sequence. N denotes the size of the block, and C_1, C_2 are constants according to the definition of SSIM index [11]. Only the original block is used because the distorted one cannot be accessed before the actual encoding. Moreover, to be compatible with the HEVC standard, only f_{ac} is applied to derive ΔQP for each coding unit.

The divisive normalization process transfers the perceptual importance to the transform coefficients, so that the coefficients with higher energy correspond to higher perceptual importance, and vice versa. This provides us with more flexibility in modeling the RD relationship for VBR rate control.

III. Two-Pass VBR Rate Control

The flowchart of the two-pass rate control algorithm is presented in Fig. 1. The first pass encoding is performed with a constant QP, and the statistics are recorded for the second pass. Before the second pass coding, GoP level bit allocation is carried out to derive the optimal bit assignment for each GoP, which is further adjusted by frame level and CU level rate control to achieve even better performance.

A. GoP level Bit Allocation

The optimal bit allocation for perceptual VBR coding is formulated as follows,

$$\min\{D\} \quad \textit{subject to} \sum_{i=1}^{n} R_i \le R_c, \tag{5}$$

where R_i represents the coding rate for each GoP, and R_c is the constraint on the total permissible rate.

Since the ultimate receiver of video is the Human Visual System (HVS), the correct optimization goal should be the overall perceptual quality. Existing rate control algorithms typically optimize the sum of absolute difference (SAD) or mean square error (MSE) with the constraint of frame level quality smoothness. However, it is widely recognized that maintaining a frame level constant MSE does not ensure constant perceived video quality. In this work, the overall quality of the whole video is defined as the average distortion in terms of the SSIM-based divisive normalized MSE for each GoP,

$$D = \sum_{i=1}^{n} D_i, \tag{6}$$

where D_i denotes the MSE in the divisive normalization domain for the i^{th} GoP. According to the divisive normalization process, the GoPs with the same MSE in the pixel domain may produce different D_i, because smaller normalization factors are assigned to perceptually more important GoPs.

Assume the Lagrange multiplier of the i^{th} GoP is λ_i, the optimal strategy that can achieve the minimization of D is to maintain a constant level of λ_i for all GoPs [18],

$$\lambda_1 = \lambda_2 = ... = \lambda_i = \lambda_n. \tag{7}$$

To find the optimal λ value, we start with an initial guess and iteratively adjust it until the best λ^* is obtained for $\sum_{i=1}^{n} R_i(\lambda^*) = R_c$, within a convex hull approximation [19].

It is noted that λ here is not the one specified by the encoder. For example, in HM codec λ is only determined by the frame type, QP and frame level, regardless of the properties of the video content. In view of various video content, the λ derivation should be adapted to the properties of the input sequences (statistical properties of residuals, structural information, etc.) [12]. For the same QP value with different residual energy, the optimal λ spans a wide range [16].

Theoretically, the optimal λ is obtained by calculating the derivative of the rate distortion cost J with respect to R, then setting it to zero, which is formulated as

$$\frac{dJ}{dR} = \frac{d(D + \lambda R)}{dR} = \frac{dD}{dR} + \lambda = 0, \tag{8}$$

leading to

$$\lambda = -\frac{dD}{dR}. \tag{9}$$

To derive λ, statistical models of both rate and distortion should be established. We employ the Laplacian distribution, which achieves a good balance between complexity and accuracy. The density of the normalized transformed residuals x in the Laplace distribution is given by

$$f_{Lap}(x) = \frac{\Lambda}{2} \cdot e^{-\Lambda \cdot |x|}, \tag{10}$$

where Λ denotes the Laplacian distribution parameter.

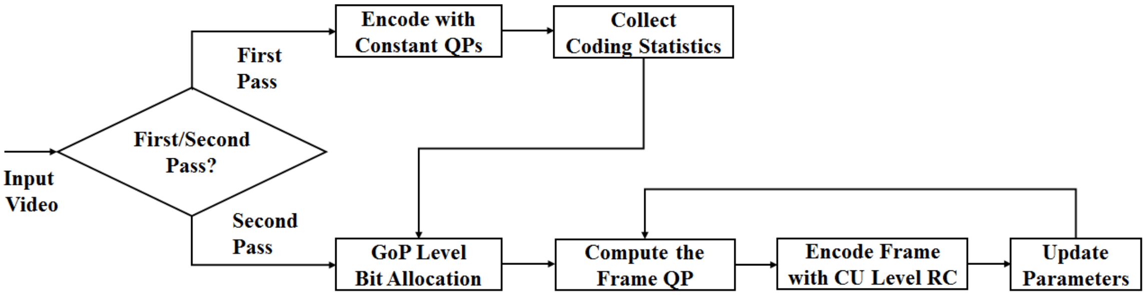


Fig. 1. Flowchart of the proposed two-pass rate control algorithm.

Considering the quantization process with quantization step Q and rounding offset γ, the distortion and rate can be modeled as [20],

$$\begin{aligned}D_i =&\alpha \cdot \left(\int_{-(Q-\gamma Q)}^{(Q-\gamma Q)} x_i^2 f_{Lap}(x_i)dx_i + \right.\\ &\left. 2\sum_{n=1}^{\infty} \int_{nQ-\gamma Q}^{(n+1)Q-\gamma Q} (x_i - nQ)^2 f_{Lap}(x_i)dx_i \right) \quad (11)\\ R =&\beta \cdot \left(-P_0 \cdot \log_2 P_0 - 2\sum_{n=1}^{\infty} P_n \cdot \log_2 P_n \right),\end{aligned}$$

where α and β are control parameters to ensure the accuracy of the estimation. In (11), the probabilities of the transformed residuals that are quantized to the zero-th and n-th quantization levels P_0 and P_n are modeled by the Laplace distribution as well.

Before the second pass encoding, the λ-Q curve for each GoP is obtained by incorporating (11) into (9). This implies that due to the divisive normalization process, the derivation of λ spontaneously takes the perceptual factors into consideration by calculating λ with residual distribution in perceptually uniform domain. For example, there are two GoPs with the same prediction residual distribution but different perceptual importance, in which the first GoP is more important with a smaller normalization factor f. Assume the Lagrange multipliers of the two GoPs are λ_1 and λ_2, respectively. The first GoP has relatively smaller Λ, so that for the same Q, $\lambda_1 > \lambda_2$. This indicates that to achieve a balance between the two GoPs, it is reasonable to lower λ_1 by borrowing more bits from the second GoP to the first GoP, so that λ_2 will increase until $\lambda_1 = \lambda_2$. Otherwise, it is always beneficial to perform bit allocation to achieve better overall quality. This is also the case when the two GoPs have the same perceptual importance but different prediction residual energy.

B. Frame and CTU level Rate Control

The task of the frame level rate control is to derive an appropriate QP value according to the target bits. Though (11) provides a solution in modeling the R-Q relationship, it is difficult to directly compute QP from the input R. Motivated by the RD analysis in HEVC [4], [21] and TM5 [22], we apply the sum of absolute transformed differences (SATD) in divisive normalization domain for QP derivation, which is formulated as

$$R = \alpha \cdot X/QP, \quad (12)$$

where X denotes the relative complexity computed by

$$X = (\frac{\sum_{i=0}^{n} w_i \cdot DN\_SATD_i}{\sum_{i=0}^{n-1} w_i \cdot DN\_SATD_i})^{\beta} \cdot R_{n-1} \cdot QP_{n-1}, \quad (13)$$

where DN_SATD_i denotes the SATD in the divisive normalization domain and w_i represents the relative weighting for each frame:

$$w_i = 0.5^{n-1} / \sum_{i=0}^{n} 0.5^{n-i}. \quad (14)$$

The parameter DN_SATD_i estimates the perceptual complexity at the frame level by computing the SATD in the divisive normalization domain, which implies that perceptually important frame will consume more bits because of the energy amplification of residuals. In our implementation, to compute Q before coding the frame, LCU level motion estimation is performed to compute DN_SATD_i.

The CU level rate control is performed by dynamically assigning each CU an appropriate ΔQP value according to its relative importance. As the frame level coding bits is derived in the perceptually uniform domain, it becomes natural to perform divisive normalization for each CU with (1), where the expectation $E(\cdot)$ in (5) is obtained within each frame to compute the relative importance of each CU. This provides the foundation of the proposed rate control algorithm, such that the optimization in GoP and frame level are both achieved in the divisive normalization domain.

IV. Experimental Results

To verify the efficiency of the proposed rate control scheme, we integrate it into the HM13.0 software [23] and compare it with the recommended algorithm (both frame and LCU levels) in HM [5]. The RD performance and rate control accuracy is evaluated with multiple scene video sequences of different resolutions in random access main configuration (RA_Main) and low delay B main configuration (LDB_Main). The rate control accuracy is measured by

$$Accu = \frac{|R_{target} - R_{actual}|}{R_{target}} \times 100\% \quad (15)$$

TABLE I
PERFORMANCE COMPARISON BASED ON THE R-λ METHOD [5] (RA_MAIN).

Sequence (Seq1∼Seq7)	R_{target}	Anchor				Proposed				ΔR^*	ΔR^{**}
		R_{actual}	SSIM	MS-SSIM	$BitErr$	R_{actual}	SSIM	MS-SSIM	$BitErr$		
Basketballpass@WQVGA	756.67	757.57	0.9486	0.9913	0.12%	757.51	0.9527	0.9922	0.11%	-6.61%	-5.53%
BlowingBubbles@WQVGA	366.76	367.53	0.9039	0.9802	0.21%	367.66	0.9099	0.9817	0.25%		
BQSquare@WQVGA	180.46	181.07	0.8373	0.9581	0.34%	181.13	0.8449	0.9607	0.37%		
RaceHorses@WQVGA	90.63	91.02	0.7579	0.9222	0.43%	90.78	0.7615	0.9227	0.17%		
Coastguard@CIF	563.72	557.47	0.9644	0.9927	1.11%	563.27	0.9689	0.9940	0.08%	-16.0%	-20.13%
Container@CIF	272.79	272.70	0.9349	0.9852	0.03%	273.63	0.9430	0.9876	0.31%		
Flower@CIF	134.31	134.60	0.8893	0.9711	0.21%	135.05	0.9028	0.9765	0.55%		
News@CIF	68.39	68.65	0.8223	0.9407	0.38%	68.91	0.8445	0.9570	0.77%		
Flowervase@WVGA	1379.94	1386.57	0.9352	0.9761	0.48%	1379.14	0.9527	0.9879	0.06%	-48.0%	-63.66%
Keiba@WVGA	642.00	648.76	0.9055	0.9550	1.05%	643.59	0.9232	0.9727	0.25%		
Mobisode@WVGA	314.82	347.26	0.8876	0.9446	10.31%	315.76	0.9136	0.9738	0.30%		
RaceHorses@WVGA	156.15	178.95	0.8577	0.9238	14.60%	156.28	0.8678	0.9495	0.09%		
Mobcal@720P	11186.72	11186.81	0.9306	0.9886	0.00%	11190.06	0.9405	0.9911	0.03%	-29.9%	-42.96%
Parkrun@720P	4822.68	4822.69	0.9000	0.9800	0.00%	4827.57	0.9190	0.9866	0.10%		
Shields@720P	2179.14	2179.15	0.8551	0.9635	0.00%	2192.39	0.8788	0.9763	0.61%		
	974.10	967.76	0.7869	0.9353	0.65%	978.86	0.8070	0.9533	0.49%		
BigShip@720P	3002.67	3002.67	0.9583	0.9902	0.00%	3002.24	0.9616	0.9912	0.01%	-19.0%	-25.31%
Raven@720P	1283.18	1283.18	0.9368	0.9811	0.00%	1285.64	0.9427	0.9840	0.19%		
ShuttleStart@720P	584.52	584.84	0.9018	0.9603	0.05%	587.07	0.9132	0.9700	0.44%		
	271.36	271.57	0.8619	0.9316	0.07%	273.03	0.8722	0.9438	0.61%		
Sunflower@1080P	3706.52	3740.76	0.9539	0.9900	0.92%	3723.05	0.9572	0.9912	0.45%	-11.9%	-11.24%
Tractor@1080P	1744.67	1764.65	0.9345	0.9816	1.15%	1753.63	0.9396	0.9839	0.51%		
Kimono@1080P	863.28	873.39	0.9063	0.9669	1.17%	865.61	0.9113	0.9696	0.27%		
ParkScene@1080P	445.18	449.62	0.8682	0.9411	1.00%	447.47	0.8729	0.9452	0.51%		
Cactus@1080P	16504.51	16504.60	0.9164	0.9858	0.00%	16503.10	0.9240	0.9882	0.01%	-20.9%	-23.41%
BasketballDrive@1080P	7623.10	7639.32	0.8855	0.9756	0.21%	7622.85	0.8996	0.9812	0.00%		
Crowd_run@1080P	3731.18	3745.64	0.8445	0.9581	0.39%	3727.32	0.8588	0.9658	0.10%		
	1860.52	1873.15	0.7954	0.9309	0.68%	1860.38	0.8035	0.9375	0.01%		

* Rate reduction while maintaining SSIM.
** Rate reduction while maintaining MS-SSIM.

TABLE II
PERFORMANCE COMPARISON BASED ON THE R-λ METHOD [5] (LDB_MAIN).

Sequences (Seq1∼Seq7)	R_{target}	Anchor				Proposed				ΔR^*	ΔR^{**}
		R_{actual}	SSIM	MS-SSIM	$BitErr$	R_{actual}	SSIM	MS-SSIM	$BitErr$		
Basketballpass@WQVGA	859.37	859.18	0.9461	0.9910	0.02%	859.19	0.9524	0.9926	0.02%	-9.25%	-11.34%
BlowingBubbles@WQVGA	393.53	393.45	0.8965	0.9786	0.02%	393.40	0.9053	0.9819	0.03%		
BQSquare@WQVGA	180.46	180.42	0.8222	0.9530	0.02%	180.43	0.8319	0.9579	0.01%		
RaceHorses@WQVGA	85.75	85.74	0.7404	0.9132	0.00%	85.77	0.7475	0.9180	0.03%		
Coastguard@CIF	642.42	642.37	0.9618	0.9919	0.01%	643.95	0.9580	0.9908	0.24%	-15.6%	-19.36%
Container@CIF	283.81	283.81	0.9281	0.9830	0.00%	283.73	0.9367	0.9857	0.03%		
Flower@CIF	126.44	126.44	0.8789	0.9678	0.00%	126.37	0.8920	0.9730	0.05%		
News@CIF	55.31	55.31	0.8066	0.9409	0.00%	55.30	0.8234	0.9494	0.02%		
Flowervase@WVGA	1527.82	1527.41	0.9384	0.9774	0.03%	1528.00	0.9547	0.9896	0.01%	-55.7%	-73.96%
Keiba@WVGA	664.59	664.42	0.9048	0.9564	0.03%	664.54	0.9414	0.9857	0.01%		
Mobisode@WVGA	304.07	304.07	0.8729	0.9279	0.00%	303.69	0.9085	0.9721	0.12%		
RaceHorses@WVGA	144.30	144.40	0.8436	0.9019	0.07%	144.28	0.8642	0.9465	0.02%		
Mobcal@720P	12618.54	12618.61	0.9299	0.9882	0.00%	12615.32	0.9390	0.9907	0.03%	-41.4%	-52.58%
Parkrun@720P	5116.98	5117.10	0.8942	0.9778	0.00%	5115.71	0.9080	0.9825	0.02%		
Shields@720P	2050.96	2051.03	0.8476	0.9593	0.00%	2050.79	0.8836	0.977	0.01%		
	810.00	805.72	0.7882	0.9297	0.53%	811.18	0.8142	0.9544	0.15%		
BigShip@720P	3141.34	3141.43	0.9570	0.9897	0.00%	3141.96	0.9605	0.9908	0.02%	-13.9%	-10.09%
Raven@720P	1172.40	1172.02	0.9346	0.9809	0.03%	1172.51	0.9392	0.9824	0.01%		
ShuttleStart@720P	464.30	463.57	0.8988	0.9636	0.16%	464.04	0.9051	0.9661	0.06%		
	185.17	185.17	0.8489	0.9293	0.00%	185.16	0.8578	0.9341	0.00%		
Sunflower@1080P	3854.38	3854.26	0.9521	0.9897	0.00%	3852.33	0.9561	0.9912	0.05%	-16.9%	-14.40%
Tractor@1080P	1711.08	1706.49	0.9293	0.9798	0.27%	1712.80	0.9359	0.9824	0.10%		
Kimono@1080P	793.85	790.10	0.8953	0.9612	0.47%	795.63	0.9050	0.9661	0.22%		
ParkScene@1080P	383.60	384.94	0.8536	0.9295	0.35%	383.89	0.8632	0.9369	0.08%		
Cactus@1080P	18201.93	18201.98	0.9171	0.9863	0.00%	18202.30	0.9204	0.9876	0.00%	-20.1%	-20.93%
BasketballDrive@1080P	8125.05	8125.07	0.8831	0.9756	0.00%	8127.74	0.8985	0.9811	0.03%		
Crowd_run@1080P	3850.69	3850.71	0.8404	0.9572	0.00%	3853.43	0.8547	0.9644	0.07%		
	1863.99	1863.99	0.7898	0.9282	0.00%	1864.80	0.7956	0.9333	0.04%		

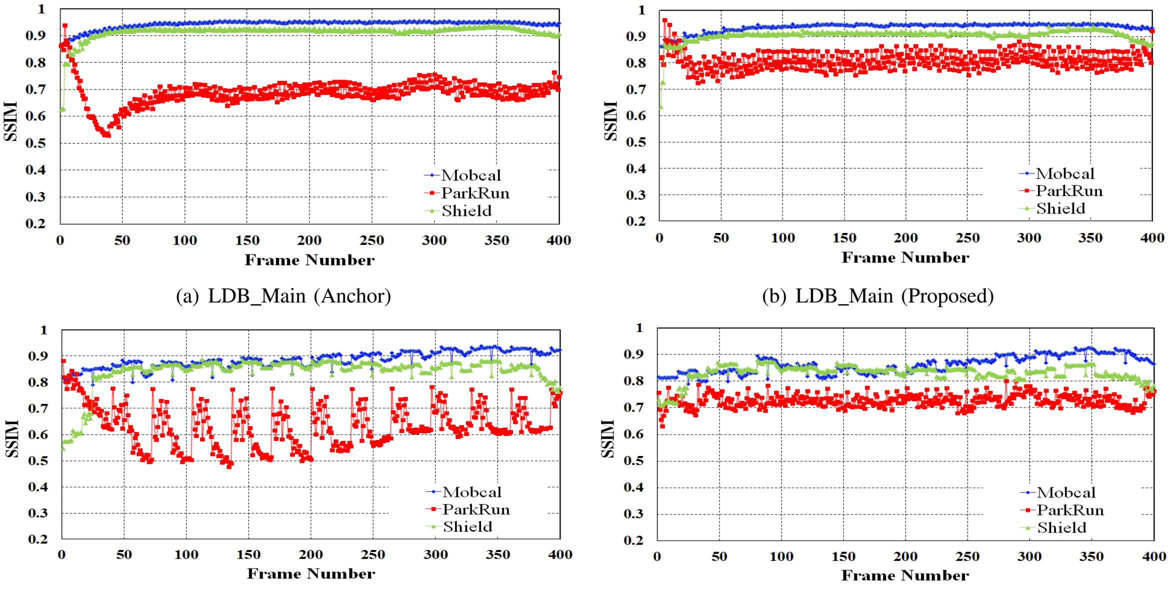
* Rate reduction while maintaining SSIM.
** Rate reduction while maintaining MS-SSIM.

The rate control performance in terms of the BD-Rate and control accuracy is demonstrated in Table I&II. Each test video is generated by three or four video shots with different statistical properties. It can be observed that the proposed bit allocation scheme can significantly improve the rate distortion performance. On average, in terms of SSIM, 24.7% bit rate reduction for LDB_Main and 21.7% bit rate reduction for RA_Main are observed. This is because the proposed bit allocation method ensures a global optimal bit distribution according to the statistics of the first pass encoding. Moreover, the proposed scheme is also able to achieve high control accuracy, which enables its applications in real scenarios.

We further demonstrate the SSIM variations for one 720P sequences in Fig. 2. To quantitatively evaluate the variations, the standard deviations of SSIM of the anchor and proposed scheme are computed as well. We can observe that although our approach does not involve a smooth term in the quality evaluation, it can create smoother quality sequences. One can discern that more bits are allocated into sequence $Parkrun$ from $Mobcal$ and $Shield$, so that the quality of the reconstructed video is much smoother with low SSIM variance. This originates from the divisive normalization based rate control approach, which automatically allocates more bits to the areas that may create more perceptual distortion, and therefore results in smoother video quality.

V. CONCLUSIONS

In this paper, we propose a two-pass perceptual VBR coding scheme based on an SSIM inspired divisive normalization framework. The novelty of the scheme lies in hierarchically constructing a perceptually uniform space for GoP, frame, and CU level rate control. The λ equalization with Laplacian distribution modeling of the transform residuals is employed, which adaptively allocates the coding bits to each GoP. R-



(a) LDB_Main (Anchor)

(b) LDB_Main (Proposed)

(c) RA_Main (Anchor)

(d) RA_Main (Proposed)

Fig. 2. Frame level SSIM comparison between anchor (left) and the proposed scheme (right) on LDB_Main (top) and RA_Main (bottom) for seq4. The standard deviations of SSIM are: (a) 0.1198; (b) 0.059; (c) 0.126; (d) 0.065.

Q relationship based on SATD in the divisive normalization domain is further adopted to obtain the frame level QP given a target bit rate. The proposed scheme demonstrates high rate control performance in terms of both control accuracy and RD performance.

References

[1] G. J. Sullivan, J. Ohm, W.-J. 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.

[2] I.-K. Kim, J. Min, T. Lee, W.-J. Han, and J. Park, "Block partitioning structure in the hevc standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1697–1706, 2012.

[3] H. Li, B. Li, and J. Xu, "Rate-distortion optimized reference picture management for high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1844–1857, 2012.

[4] J. Si, S. Ma, S. Wang, and W. Gao, "Laplace distribution based CTU level rate control for HEVC," in *Visual Communications and Image Processing (VCIP), 2013*, 2013, pp. 1–6.

[5] B. Li, H. Li, L. Li, and J. Zhang, "Rate control by r-lambda model for HEVC," in *JCTVC-K0103, JCTVC of ISO/IEC and ITU-T, 11th meeting Shanghai, China*, 2012.

[6] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Rate-GOP based rate control for high efficiency video coding," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1101–1111, 2013.

[7] S. Wang, S. Ma, L. Zhang, S. Wang, D. Zhao, and W. Gao, "Multi layer based rate control algorithm for HEVC," in *IEEE International Symposium on Circuits and Systems*, 2013, pp. 41–44.

[8] W.-N. Lie, C.-F. Chen, and T. C.-I. Lin, "Two-pass rate-distortion optimized rate control technique for H.264/AVC video," in *Visual Communications and Image Processing*. International Society for Optics and Photonics, 2005, pp. 596 035–596 035.

[9] J. Sun, Y. Duan, J. Li, J. Liu, and Z. Guo, "Rate-distortion analysis of dead-zone plus uniform threshold scalar quantization and its applicationpart II: Two-pass VBR coding for H.264/AVC," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 215–228, 2013.

[10] D. Zhang, K. N. Ngan, and Z. Chen, "A two-pass rate control algorithm for H.264/AVC high definition video coding," *Signal Processing: Image Communication*, vol. 24, no. 5, pp. 357–367, 2009.

[11] 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.

[12] 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.

[13] 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 Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 11, pp. 1614–1624, 2010.

[14] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Rate-ssim optimization for video coding," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 833–836.

[15] D. Brunet, E. R. Vrscay, and Z. Wang, "On the mathematical properties of the structural similarity index," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1488–1499, 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. Rehman and Z. Wang, "SSIM-inspired perceptual video coding for HEVC," in *IEEE International Conference on Multimedia and Expo*, 2012, pp. 497–502.

[18] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and mpeg video coders," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 533–545, 1994.

[19] A. Ortega, "Optimization techniques for adaptive quantization of image and video under delay constraints," Ph.D. dissertation, Columbia University, 1994.

[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] S. Ma, J. Si, and S. Wang, "A study on the rate distortion modeling for high efficiency video coding," in *IEEE International Conference on Image Processing*. IEEE, 2012, pp. 181–184.

[22] MPEG, "TM5," http://www.mpeg.org/MPEG/MSSG/tm5.

[23] K. McCann, B. Bross, W.-J. Han, I. K. Kim, K. Sugimoto, and G. J. Sullivan, "High efficiency video coding (HEVC) test model 13 (HM 13) encoder description," *JCTVC-O1002*, 2013.