PAPER

A Rate Perceptual-Distortion Optimized Video Coding HEVC

Bumshik LEE[†] and Jae Young CHOI^{††a)}, Members

SUMMARY In this paper, a perceptual distortion based rate-distortion optimized video coding scheme for High Efficiency Video Coding (HEVC) is proposed. Structural Similarity Index (SSIM) in transform domain, which is known as distortion metric to better reflect human's perception, is derived for the perceptual distortion model to be applied for hierarchical coding block structure of HEVC. A SSIM-quantization model is proposed using the properties of DCT and high resolution quantization assumption. The SSIM model is obtained as the sum of SSIM in each Coding Unit (CU) depth of HEVC, which precisely predict SSIM values for the hierarchical quadtree structure of CU in HEVC. The rate model is derived from the entropy, based on Laplacian distributions of transform residual coefficients and is jointly combined with the SSIM-based distortion model for rate-distortion optimization in an HEVC video codec and can be compliantly applied to HEVC. The experimental results demonstrate that the proposed method achieves 8.1% and 4.0% average bit rate reductions in rate-SSIM performance for low-delay and random access configurations respectively, outperforming other existing methods. The proposed method provides better visual quality than the conventional mean square error (MSE)-based RDO coding scheme.

key words: HEVC, SSIM, perceptual video coding

1. Introduction

The compressed bit streams are delivered to the end-users with the minimum distortion given a target bit rate. Since the final consumers of encoded video streams are usually human beings, the measurement of encoded video quality should strongly correlate with human perception. However, it has been known that it is practically hard to quantify the video quality in the way that the human perceives. In video compression, video encoding generally performs rate-distortion optimization (RDO) to choose the optimum block sizes of motion estimation/compensation and transform, which affects the perceptual quality in compressed video. In this case, it is desirable to appropriately measure the distortion that can best reflect the perceptual visual quality. The objective distortion metrics such as sum of squared errors (SSE) and sum of absolute differences (SAD) have been widely used for RDO in video codecs.

However, even though these objective quality metrics are computationally beneficial and easy to model, they do not well reflect video quality that human perceives [1].

Recently, many video quality metrics that are well correlated with human visual system (HVS) have been developed. Structural Similarity Index (SSIM) utilizes the property that structural distortion significantly affects the perceptual image quality degradation. It has been widely used as a perceptual quality metric in many video applications.

High Efficiency Video Coding (HEVC) [2] standard has officially been approved as ITU-T Recommendation H.265 and ISO/IEC 23008-2 (MPEG-H Part 2). The HEVC has been developed with the objective of achieving the coding efficiency improvement of about 50% or more, compared to the previous video coding standard such as H.264/AVC [3]. Owing to the expanded prediction and transform block sizes with a flexible coding structure, HEVC can efficiently encode the video sequences from low to high picture resolutions with various signal characteristics. Figure 1 illustrates an example of hierarchical block partition structures in HEVC. As shown in Fig. 1, flexible block structures in HEVC are adopted with quadtree partitions and higher depth levels, which are composed of Coding Unit (CU), Prediction Unit (PU) and Transform Unit (TU) [4], [5]. The CU is a basic processing unit for encoding and decoding, which includes motion estimation (ME) and motion compensation (MC), transform, quantization and entropy coding etc. The CU with the maximum size is called the Coding Tree Unit (CTU) for which its size and the number of predefined depths are signaled in a sequence level. Due to such complicated block structures of variable block-sized transform kernels, various prediction block sizes for ME and MC, etc., it became difficult to model perceptual distortion of videos compressed using HEVC.

In this paper, we develop a rate distortion optimization coding scheme for HEVC where the tradeoff between the rates and the SSIM-based perceptual distortions is made in HEVC encoding process. The distortion is minimized subject to a rate constraint in RDO, which can be expressed as

$$\min\{D\} \quad \text{subject to} \quad R \leq R_c \quad (1)$$

Fig. 1 An example of hierarchical block partition structures in HEVC.

Manuscript received May 11, 2018.

Manuscript revised July 31, 2018.

Manuscript publicized August 24, 2018.

[†]The author is with Chosun University, 309 Pilmoon-Daero, Gwangju, 61452, Korea.

^{††}The author with Hankuk University of Foreign Studies (HUFS), Yongin-si, Gyeonggi-do, 17305, Korea.

a) E-mail: jychoi@hufs.ac.kr (Corresponding author)

DOI: 10.1587/transinf.2018EDP7164

where D and R are distortion and rate, respectively, and R_c is a rate constraint. In the conventional RDO, D is measured by mean square errors (MSE) or SAD. The constrained RDO problem in (1) can be solved via a unconstrained RDO problem as

$$\min\{J\} \quad \text{where} \quad J = D + \lambda \cdot R \quad (2)$$

where J is a rate-distortion cost, and λ is a Lagrange multiplier. In this problem, finding the optimal Lagrange multiplier is very important because it determines tradeoff between distortion and rate, affecting the rate-distortion performance. In general, λ is a fixed value over entire coding process and is derived from rate and distortion models. We often use $\lambda = 0.85 \times 2^{(QP-12)/3}$ for H.264/AVC and $\lambda = 0.49 \times 2^{(QP-12)/3}$ for HEVC [6]–[9] where QP is a quantization parameter (QP). In the conventional MSE-based coding scheme, more adaptive schemes have been proposed in [10]–[12] by developing rate and distortion models more elaborately for the previous coding standards which have relatively simple structures compared to HEVC.

On the other hand, much effort has been made for perceptual video coding using SSIM. Huang *et al.* [13] proposed the distortion metric using $(1 - \text{SSIM})$ for rate-distortion optimization of H.264/AVC. To calculate a Lagrange multiplier, they assumed that the tangent at a certain point of the rate-distortion curve is the same as the tangent at the closest operating point on the rate-SSIM curve for SSIM-based RDO. In [13], an approximation method was proposed; herein, a rate-distortion point of a previous coded frame is projected to the RD curve in horizontal and vertical directions. Then, a Lagrange multiplier are calculated based on two projected points on the RD curve. This method is applied for H.264/AVC perceptual rate control in [14]. However, the two schemes in [13], [14] require every key picture to be encoded twice by two different QP values in order to obtain the model parameters of the rate-SSIM curve for the current frame under encoding.

In [15], a SSIM-QP model and a rate-Q model were proposed for SSIM-based RDO encoding in H264/AVC. The proposed SSIM-QP model is based on the transform domain assuming that transform residuals follow a Laplacian probability density function (PDF) while the rate-quantization (R-Q) model was taken from the work in [10]. Then, the Lagrange multiplier is computed from the proposed two models. For the SSIM-QP model, it is computed from DCT sub-bands in a transform block that has 16 sub-bands for the 4×4 transform kernel of H.264/AVC. However, it is hard to model SSIM-based distortion if variable block transforms are used in such a case of HEVC where variable block-sized transform sizes ranging from 4×4 to 32×32 are selectively used in the RDO sense. In this case, a total of 1,360 sub-band statistics are necessarily to be collected which is not computationally efficient.

In [16], [17], perceptual video coding schemes were proposed for H.264/AVC [16] and HEVC [17]. The divisive normalization factors derived from transform domain SSIM are applied in the quantization process. Quantization

matrices for perceptual normalization and quantization are made for quantization process. Although they show significant perceptual coding gains, a number of normalization factors in each sub-band should be calculated in both encoder and decoder sides, resulting in complexity increase especially in decoder sides. Furthermore, it is not compliant with H.264/AVC and HEVC coding standards.

Yeo [18], [19] proposed a distortion metric using SSIM. Unlike most of the other methods, he used the inverse of SSIM ($1/\text{SSIM}$) instead of the method $(1 - \text{SSIM})$ for the perceptual distortion modeling. Moreover, the proposed scheme in an RDO framework does not require computing SSIM explicitly. From the relations between MSE and SSIM, the RDO is performed in a computationally efficient manner. However, the local variances of the original image are obtained in CTU levels if the Yeo's method is applied for HEVC. The local variances in CTU levels do not reflect well the statistics in their lower CU levels, thus resulting in incorrect Lagrange multiplier values in CU levels.

In [20], a SSIM-motivated rate control scheme is proposed for HEVC. A divisive normalization proposed in [16], [17] is used for perceptual distortion model in this scheme, where a variable bitrate (VBR) rate control with two passes is proposed. In the first coding pass, perceptual Lagrange multiplier using the model is obtained, then bit allocation is performed in the second pass. In [21], the relation between rate and SSIM is proposed. Although it shows good relations, it might not efficiently be used because the rate model is not a function of quantization (q).

As one of the perceptual video coding schemes, a just-noticeable difference (JND) model based perceptual video coding was proposed in [22], where a pixel-domain JND model for the transform skip mode of HEVC and a transform-domain JND model for the transform non-skip modes of HEVC and shows promising bit-rate reduction with similar subjective quality.

The aforementioned methods for perceptual video coding show significant improvements on rate-SSIM performances or subjective video quality assessments. However, some of the proposed methods cannot be applied to the current coding structure of HEVC that has hierarchical quadtree coding block structures and a quadtree-based variable block-sized transform structure. Moreover, little study has been reported for perceptual video coding that is dedicated for HEVC codecs.

The rest of the paper is organized as follows: In Sect. 2, an SSIM-based distortion metric and a rate-Q model are proposed. Then, we discuss the Lagrange multiplier with the proposed models and implementation issues; the experimental results are provided in Sect. 3. We conclude the paper in Sect. 4.

2. Proposed Method

2.1 A Proposed Lagrange Multiplier for SSIM Based Rate-Distortion Optimization (RDO)

The SSIM between a reference and a distorted image is defined in [1] and is given by

$$SSIM(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (3)$$

where $\mathbf{x}$ and $\mathbf{y}$ are two image regions to be compared, μ_x and μ_y are mean values of $\mathbf{x}$ and $\mathbf{y}$, respectively. σ_x^2 and σ_y^2 are the variances of $\mathbf{x}$ and $\mathbf{y}$, respectively. σ_{xy} is the cross correlation between $\mathbf{x}$ and $\mathbf{y}$. C_1 and C_2 are constants to avoid instability. In [18], C_1 and C_2 are recommended as $C_1 = (0.01 \times (2^{\text{bit_depth}} - 1))^2$ and $C_2 = (0.03 \times 2^{\text{bit_depth}} - 1)^2$, respectively. In this paper, we propose a DCT transform domain SSIM distortion model. The DCT of a vector $\mathbf{x} \in \mathbf{R}^N$ can be obtained as

$$X(k) = \sqrt{\frac{2}{N}} \cdot a(k) \cdot \sum_{i=0}^{N-1} \cos\left(\frac{\pi(2i+1)k}{2N}\right) x(i) \quad (4)$$

where

$$a(k) = \begin{cases} 1/\sqrt{2}, & \text{if } k = 0 \\ 0, & \text{if } k \neq 0 \end{cases}$$

Since the DCT is a unitary transform and by the Parseval theorem [16], the following relationships can be obtained

$$\mu_x = \frac{\sum_{i=0}^{N-1} x(i)}{N} = \frac{X(0)}{\sqrt{N}} \quad (5)$$

$$\sigma_x^2 = \frac{\sum_{i=0}^{N-1} x(i)^2 - N\mu_x^2}{N-1} = \frac{\sum_{k=1}^{N-1} X(k)^2}{N-1} \quad (6)$$

$$\sigma_{xy} = \frac{\sum_{i=0}^{N-1} x(i)y(i) - N\mu_x\mu_y}{N-1} = \frac{\sum_{k=1}^{N-1} X(k)Y(k)}{N-1} \quad (7)$$

In general, a reconstructed image in the hybrid codecs is modeled by adding residual error to the original one as

$$y = x + e \quad (8)$$

where x is the original image, y is the reconstructed image and e is the residual image. If x and e are uncorrelated, the variance of the reconstructed image can be expressed as

$$\sigma_y^2 = \sigma_{x+e}^2 = \sigma_x^2 + \sigma_e^2 + 2\sigma_{xe} \approx \sigma_x^2 + \sigma_e^2 \quad (9)$$

In (8), the residual term can be ignored i.e., $y - x = e \approx 0$ under the high resolution quantization assumption. In addition, we assume that μ_x and μ_y are identical in the hybrid codecs. Thus, (3) is simplified to

$$SSIM(\mathbf{x}, \mathbf{y}) \approx \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (10)$$

By substituting (5), (6) and (7) into (10), the transform domain SSIM index is obtained as

$$SSIM(\mathbf{x}, \mathbf{y}) \approx \left(\frac{2 \frac{\sum_{k=1}^{N-1} X(k)Y(k)}{N-1} + C_2}{\frac{\sum_{k=1}^{N-1} X(k)^2 + Y(k)^2}{N-1} + C_2} \right) = \left(1 - \frac{\sum_{k=1}^{N-1} (X(k) - Y(k))^2}{\sum_{k=1}^{N-1} (X(k)^2 + Y(k)^2) + (N-1)C_2} \right) \quad (11)$$

where $(X(k) - Y(k))^2$ is denoted as $D(k)$, which is a squared error, i.e., $(X(k) - Y(k))^2 = e^2(k) = D(k)$. Since the residual signal is negligibly small in the assumption, we have a relationship such as

$$Y(k)^2 + X(k)^2 - 2Y(k)X(k) = e(k)^2 \approx 0 \quad (12)$$

Therefore,

$$Y(k)^2 + X(k)^2 \approx 2Y(k)X(k) \quad (13)$$

The variance of the residual signal is obtained as

$$\text{VAR}(Y - X) = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy} = \sigma_e^2 \quad (14)$$

Then, the covariance of x and y is equal to the variance of x as

$$\sigma_{xy} = \sigma_x^2 \quad (15)$$

Hence, using (7), (13) and (15), (11) can be simplified to

$$SSIM(\mathbf{x}, \mathbf{y}) \approx 1 - \frac{D}{2\sigma_x^2 + C_2} \quad (16)$$

where σ_x^2 is a variance of pixel values in the original image, and D is the MSE between the original and distorted images and is computed as $1/(N-1) \sum_{k=1}^{N-1} (X(k) - Y(k))^2$. The MSE can be expressed as a function of QP. In our previous works [23]–[26], we observed that the transform coefficients for predicted residues in the HEVC codec follow a Laplacian PDF as

$$f_L(x) = \frac{\gamma}{2} e^{-\gamma|x|} \quad (17)$$

where γ is a model parameter of Laplacian PDF and is computed as $\sqrt{2}/\sigma$ [27], and σ is an average standard deviation of transform coefficients. Hence, D can be computed in the dead zone plus uniform threshold (DZ + UTQ) [28] as

$$D = \int_{(i-f) \cdot q}^{(i+1-f) \cdot q} (l - iq)^2 f_L(l) dl = \int_{-(q-fq)}^{q-fq} l^2 f_L(l) dl + 2 \sum_{i=1}^{\infty} \int_{(i-f) \cdot q}^{(i+1-f) \cdot q} (l - iq)^2 f_L(l) dl \quad (18)$$

where f is the rounding offset parameter for quantization, l is the random variable of Laplacian PDF and q is a quantization step size. When a uniform distribution in a quantization interval is valid in the high resolution quantization

assumption, γ approaches to zero [27]. Therefore, (18) can be simplified to

$$\begin{aligned}\hat{D} &\approx \lim_{\gamma \rightarrow 0} \frac{(2 + \gamma q - 2f\gamma q) \cdot \gamma q \cdot e^{f\gamma q} + 2 - 2e^{\gamma q}}{\gamma^2(1 - e^{\gamma q})} \\ &= \frac{q^2(3f^2 - 3f + 1)}{3}\end{aligned}\quad (19)$$

For general use, we denote (19) as $D = \alpha q^2$ where α is a constant as $\alpha = (3f^2 - 3f + 1)/3$. Then, SSIM-based distortion model can be obtained with $(1 - \text{SSIM})$ as

$$\begin{aligned}D_{\text{SSIM}} &\equiv 1 - \text{SSIM} \\ D_{\text{SSIM}} &\approx \frac{D}{2\sigma_x^2 + C_2} = \frac{\alpha q^2}{2\sigma_x^2 + C_2}\end{aligned}\quad (20)$$

As a matter of fact, there has been a similar work of the SSIM model in transform domain for H.264/AVC codec. In [15], Wang proposed the SSIM factor defined as

$$M_{RR} = \left(1 - \frac{D_0}{2\sigma_0^2 + C_1}\right) \cdot \left(1 - \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{D_i}{2\sigma_i^2 + C_2}\right) \quad (21)$$

where σ_i^2 is the standard deviation of DCT coefficient for the i -th subband and N is the block size, D_i is the MSE between original and reconstructed frames in the i -th band. The SSIM estimator is defined as $\hat{S} = \alpha + \beta M_{RR}$, where α and β are model parameters. However, this model is available for a single type transform kernel such as 4×4 transform of H.264/AVC. In this case, only 16 sub-band statistics are required to compute D_{SSIM} . It is efficiently applicable to simple transform architecture of H.264/AVC where only 4×4 transform block is used for the Baseline Profile. However, it is not appropriate for the HEVC that has various block-sized transform kernels of 4×4 , 8×8 , 16×16 , and 32×32 , and for the relatively larger transform kernels of 8×8 , 16×16 , and 32×32 sizes. Thus, if the Wang's method is applied for HEVC, a total of 1,360 sub-band statistics are necessarily to be computed, which is not computationally efficient. As another related work, a perceptual video coding scheme is proposed based on a SSIM-inspired divisive normalization for H.264/AVC video codec in [16]. The key idea is to derive the SSIM-inspired quantization matrix using signal statistics and normalization factors derived from SSIM for each DCT sub-band. Final quantization matrices are obtained by multiplying normalization factors which reflect the perceptual coding factor in DCT domain to original quantization matrices. The normalization factor is derived in [29]. The Lagrange multiplier for rate-distortion optimization is not changed during encoding but contents adaptive SSIM-inspired normalized factor is multiplied to quantization matrices in quantization process in this scheme. On the other hand, a new Lagrange multiplier is proposed for perceptual video coding using $D_{\text{SSIM-Q}}$ model in which SSIM-based distortion model is derived as a function of quantization step size using the distribution of DCT coefficient and high resolution assumption in our proposed method.

In our previous work in [23]–[26], we found that the

statistics of transform coefficients are quite different according to the CU depths. The multiple Laplacian PDFs were used to model rate and distortion taking into account different statistics of residual input in different CU depth levels. Similar to our previous works, we divide the frequency components of transform coefficients not into sub-band frequencies as [15] but into CU depth categories. Another distinction of our proposed perceptual distortion model compared to the Wang's model is that the pixel variance is finally used for the perceptual distortion model instead of the variance of transform coefficients. The computation in the pixel domain is more efficiently performed in the HEVC codec. The proposed perceptual distortion is modeled as a linear combination of SSIMs weighted by their CU regions, and can be expressed as

$$\begin{aligned}D_{\text{SSIM}}(q) &= \alpha \sum_{i=0}^{N-1} w_i D_{\text{SSIM},i} \\ &= \alpha \sum_{i=0}^{N-1} w_i \frac{q^2}{2\sigma_{x,i}^2 + C_2}\end{aligned}\quad (22)$$

where N is a total number of CU depths, i is an index of CU depth, D_i is the MSE for i -th CU depth, $\sigma_{x,i}^2$ is the pixel variance of the original image in the i -th CU depth, w_i is a weighting factor for each CU depth which can be computed as the ratio of the total image region in the i -th CU depth to the entire image region, and α is a constant value and can be ignored in this section because it will be merged with another constant value in (32). In [30], a synthesized Laplacian model instead of a mixture model is used for rate control application. Although it is claimed that the mixture model in the proposed method has difficulties in numerical solution [30], it can produce more accurate distortion estimation. It has been already investigated in the previous work in [23]. In addition, a Cauchy distribution model is used for rate and distortion modeling in [31] instead of the Laplacian distribution. A Cauchy model based RD model shows good results for intra-predicted residuals owing to its shape with long tails compared to the Laplacian model. However, as compared in [24] of our previous works, a Cauchy PDF-based model does not show better results in accuracy in general.

In order to compute an optimal Lagrange multiplier, an accurate rate model is also developed. The rate models are proposed for HEVC in [17], [19], where the single Laplacian PDF based rate model is proposed for simple computation of the optimal Lagrange multiplier. A rate model in a frame level is derived by taking into account the resulting bit amounts. For the proposed rate model, the entropy for the rate is used such as

$$R(q) = \beta H(q, \gamma) \quad (23)$$

where $H(q, \gamma)$ is the entropy of quantized transform coefficients given q and γ . β is a model scaling factor and can be ignored because it will be merged with another constant value in (32). The entropy for the quantized transform coefficients is given by

$$H(q, \gamma) = -P_0 \cdot \log_2 P_0 - 2 \sum_{i=1}^{\infty} P_i \cdot \log_2 P_i \quad (24)$$

where

$$P_0 = \int_{-(q-fq)}^{q-fq} f_L(l) dl = 1 - e^{-(1-f)q\gamma} \quad (25)$$

and

$$P_i = \int_{(i-f)q}^{(i+1-f)q} f(l) dl = \begin{cases} 1/2 \cdot e^{-(i-f)q\gamma} \cdot (1 - e^{-q\gamma}), & \text{for } i > 0 \\ 1/2 \cdot e^{(i-f)q\gamma} \cdot (1 - e^{-q\gamma}), & \text{for } i < 0 \end{cases} \quad (26)$$

where P_0 and P_i are the probabilities that transform coefficients are quantized to zero and the i -th quantization interval, respectively. Note that P_0 and P_i are obtained only for non-SKIP blocks. The entropy can be expressed in a closed form by substituting (24) and (25) into (23) to obtain

$$\begin{aligned} H(q, \gamma) &= -P_0 \cdot \log_2 P_0 - 2 \sum_{i=1}^{\infty} P_i \cdot \log_2 P_i \\ &= -(1 - e^{-(1-f)q\gamma}) \cdot \log_2(1 - e^{-(1-f)q\gamma}) \\ &\quad - e^{-(1-f)q\gamma} \cdot \left[\log_2(1 - e^{-q\gamma}) - 1 + \frac{q\gamma}{\ln 2} \left(f - \frac{1}{1 - e^{-q\gamma}} \right) \right] \end{aligned} \quad (27)$$

Using (24), $H(q, \gamma)$ can be simply expressed in terms of P_0 as

$$\begin{aligned} H(q, \gamma) &= \left\{ \underbrace{\log_2(1 - e^{-(1-f)q\gamma})}_{\equiv A} - \underbrace{\log_2(1 - e^{-q\gamma})}_{\equiv B} + \underbrace{\left(\frac{2 - f + (f-1)e^{-q\gamma}}{1 - e^{-q\gamma}} \frac{q\gamma}{\ln 2} \right)}_{\equiv C} \right\} \\ &\quad \times \underbrace{e^{-(1-f)q\gamma} - \log_2(1 - e^{-(1-f)q\gamma})}_{\equiv D} \\ &= A(1 - P_0) - B(1 - P_0) + C(1 - P_0) + D \end{aligned} \quad (28)$$

where A , B and D approach zero when $q\gamma$ takes on a sufficiently large value. In this case, (28) can be simplified to

$$H(q, \gamma) \approx C \cdot (1 - P_0) \quad (29)$$

where C is a constant value. P_0 can be represented using the Laplacian distribution in (25). Therefore, the entropy for a frame can be simplified to

$$\begin{aligned} H(q, \gamma) &\approx C \cdot (1 - P_0) \\ &= C \cdot \int_{-(q-fq)}^{q-fq} f(l) dl = C \cdot e^{-(1-f)q\gamma} \end{aligned} \quad (30)$$

where $f(l)$ indicates the Laplacian PDF as (16). For generality, the proposed rate model can be represented as

$$R(q) \approx \beta e^{-(1-f)q\gamma} \quad (31)$$

We have the closed forms of the perceptual distortion model in (22) and the rate model in (31), respectively, both of which are expressed as quantization step size q . From the proposed SSIM-distortion model in (22) and proposed rate model in (31), a Lagrange multiplier can be calculated by

$$\begin{aligned} \lambda_{SSIM} &= -\frac{dD_{SSIM}/dq}{dR/dq} \\ &= \frac{\theta}{(1-f)\gamma e^{-(1-f)q\gamma}} \sum_{i=0}^{N-1} w_i \frac{2q}{2\sigma_{x,i}^2 + C_2} \end{aligned} \quad (32)$$

where θ is a constant.

2.2 Avoiding Error Propagations and Coding Details

We derived a perceptual-based Lagrange multiplier for rate distortion optimization for HEVC. However, the proposed rate-distortion optimization method is based on the assumption that the transform coefficients are modelled with Laplacian distribution and the model is static without abrupt scene changes between consecutive frames. Hence, if the assumption is not valid in actual video coding, it may suffer from the degradation of coding performance. It is necessary to maintain the sufficient quality of video frames by detecting the abrupt change of frames in terms of frame statistics. Otherwise, the Laplacian parameter γ is going to be an improper value, thus resulting in degraded rate distortion performances. To detect the abrupt changes between frames, we use the variance of the frame difference. Let denote $\sigma_{r,n}^2$ as variance of the frame residue for the n -th frame as

$$\sigma_{r,n}^2 = \frac{1}{W \times H} \sum_{x,y} [f_n(x, y) - \hat{f}_n(x, y)]^2 \quad (33)$$

where W and H are the width and height of the frame, respectively. $f_n(x, y)$ and $\hat{f}_n(x, y)$ are the n -th original and reconstructed frames, respectively. The ratio of two variances (R_n) for two successive frame difference are computed as

$$R_n = \frac{\sigma_{r,n-1}^2}{\sigma_{r,n}^2} \quad (34)$$

If R_n is satisfied with a certain condition, it is regarded as the abrupt change. In the proposed method, we set the Lagrange multiplier as

$$\lambda_{SSIM,n} = \begin{cases} \text{clip}(\xi_l \lambda_{SSIM,n-1}, \lambda_{SSIM,n}, \xi_u \lambda_{SSIM,n-1}), & \text{if } R_n > \xi_u \text{ or } R_n < \xi_l \\ (32), & \text{otherwise} \end{cases} \quad (35)$$

where ξ_u and ξ_l are threshold values which are empirically set to 1.3 and 0.7, respectively.

In order to obtain λ_{SSIM} from the rate and perceptual distortion model for the next frame encoding, a weighing factor w in (21) for each CU depth and Laplacian model

parameter γ in (30) should be determined *a priori*. In our proposed method, they can be computed using a weighted moving average method from a few previous frames in coding orders as follows:

$$\hat{w}_{n+1} = \frac{1}{W \times H} \sum_{f=0}^{N_f-1} \eta_f S_{i,n-f} \text{ and } \hat{\gamma}_{n+1} = \sum_{f=0}^{N_f-1} \eta_f \gamma_{n-f} \quad (36)$$

where $\hat{\lambda}_{n+1}$ are the model parameter estimate for the Laplacian PDF and the estimate for total number of pixels to be applied for $(n+1)$ -th frame, W and H are frame width and height, respectively. N_f is the number of the most recent frames to estimate $\hat{\lambda}_{n+1}$ and $\hat{w}_{n+1}$, which is empirically set to 3 with $\eta_0 = 0.533$, $\eta_1 = 0.333$ and $\eta_2 = 0.134$ for all the test sequences in this paper. $S_{i,n-f}$ is the total block size that the block of CU depth is selected in the $(n-f)$ -th frame, and is computed as the sum of the block sizes (pixels) of non-SKIP mode.

2.3 Overall Method

The Lagrange multiplier λ_{SSIM} for rate-perceptual distortion optimization is updated at frame level in the proposed method. For the distortion model, pixel variances shown in (22) for each CU depth level are collected in the end of encoding at a frame. On the other hand, the rate is estimated in a frame level using transform coefficients using (31) from all transform block sizes. In the next step, a new Lagrange multiplier λ_{SSIM} is calculated for the next frame coding using (32). With new λ_{SSIM} and SSIM values in a CU block, rate-perceptual distortion optimization is performed. The proposed rate-perceptual distortion optimization method is summarized in Algorithm 1:

Algorithm 1: Proposed rate-perceptual distortion optimization algorithm

- Step 1: Obtain the pixel variance and block sizes in each CU depth level for the distortion model (22)
- Step 2: Obtain the Laplacian model parameters using transform coefficients values from all TU blocks to estimate rate in a frame level as (31)
- Step 3: Model parameters and weighing factors are computed using three consecutive previous frames
- Step 3: A new SSIM Lagrange multiplier λ_{SSIM} is computed using (32). New λ_{SSIM} is applied for rate-distortion optimization for next frames
- Step 4: A SSIM value for a CU block is obtained using non-overlapped 8×8 blocks
- Step 5: Perform RDO using $D_{SSIM} + \lambda_{SSIM} \cdot R$ to decide the optimal CU size in term of perceptual distortion

In Step 4, block SSIM values are calculated. A SSIM value of a larger block in upper CU level is compared with SSIM value of 4 partitioned smaller blocks in lower CU level. In order to compute the SSIM of the block, the block is partitioned into non-overlapped 8×8 blocks and SSIM is computed in each sub-block. Thus, it is noted that the proposed SSIM-based Lagrange multiplier is used for only CU size determination. Thus, the proposed method is not used for PU and TU, where the smallest block size is smaller than 8×8 .

3. Experimental Results

First, in order to investigate the estimation accuracy of the SSIM values in the proposed method, we measured the correlation coefficient values between the actual SSIM values and the estimated SSIM. Figure 2 shows the relation between the actual SSIM values and the estimated SSIM values generated by the proposed perceptual distortion model. The experiment were performed with HM16.0 for test sequences (*BQSquare*, *BlowingBubbles*, *PartyScene*, *BasketballDrill*, *BQTerrace* and *ParkScene*) of spatial resolutions 416×240 , 832×480 and 1920×1080 . Figure 2-(a) shows the correlation between actual SSIM values and proposed perceptual model. The values are taken from all test sequences. Figure 2-(b) shows the correlation according to the spatial resolutions where both of the actual SSIM values and the estimated SSIM values by the proposed model are

Fig. 2 Relation between the actual SSIM values and the estimated SSIM values produced by the proposed SSIM model. (a) Proposed SSIM vs. actual SSIM. Correlation = 0.996; (b) SSIM $\times$ W $\times$ H for each spatial resolution. Correlation = 0.999. (416×240 : 0.996. 832×480 : 0.997, 1920×1080 : 0.997)

Table 1 BDBR (%) using SSIM

Sequences		Low Delay				Random Access			
		Huang's [13]	Rehman's [17]	Yeo's [19]	Proposed	Huang's [13]	Rehman's [17]	Yeo's [19]	Proposed
416×240	<i>BQSquare</i>	11.3	-13.6	8.8	-11.3	12.3	-7.3	11.1	-10.2
	<i>BlowingBubbles</i>	12.0	-0.1	16	-4.6	13.4	-0.5	16.5	-3.8
	<i>BasketballPass</i>	14.1	-8.2	9.4	-11.2	15.2	-2.6	10.4	-5.3
	<i>RaceHorses</i>	1.6	-6.2	2.6	-8.3	4.6	-4.5	4.7	-6.2
832×480	<i>PartyScene</i>	9.1	-6.2	7.1	-10.3	15.6	-5.0	9.2	-4.1
	<i>BasketballDrill</i>	12.9	-8.2	13.8	-11.2	14.6	-1.3	19.2	-5.9
	<i>BQMall</i>	20.2	-2.7	10.4	-8.8	25.3	-0.6	17.4	-5.3
	<i>RaceHorses</i>	2.1	-7.4	0.2	-9.9	5.9	-2.6	5.2	-4.2
1280×720	<i>Vidyo1</i>	-5.3	-3.3	-2.1	-0.9	1.3	-1.5	0.7	-0.2
	<i>Vidyo3</i>	-6.3	-4.6	-3.6	-1.6	2.3	-1.6	1.2	-0.8
	<i>Vidyo4</i>	-2.3	-6.5	-2.1	-1.5	6.9	-4.8	3.6	1.2
	<i>BQTerrace</i>	9.1	-1.5	18.8	-12.3	12.6	-0.8	24.2	-4.2
1920×1080	<i>ParkScene</i>	16.2	-2.7	14.8	-4.8	20.5	-1.1	16.4	-2.1
	<i>Kimono1</i>	8.8	-0.5	10.9	-9.6	14.5	0.0	15.2	-1.3
	<i>BasketballDrive</i>	2.2	-3.0	3.9	-15.2	20.4	-1.5	4.1	-5.3
	<i>Cactus</i>	1.2	-10.7	0.6	-9.2	5.6	-3.8	2.3	-3.1
2560×1600	<i>Traffic</i>	2.3	-9.6	0.2	-8.5	4.9	-2.5	4.2	-3.1
	<i>PeopleOnStreet</i>	0.6	-5.6	1.4	-6.9	3.6	-4.4	1.8	-2.1
Average		6.1	-5.6	6.2	-8.1	11.8	-2.6	9.3	-4.0

(a) *BQSquare* (416×240) (b) *BasketballPass* (416×240) (c) *PartyScene* (832×480) (d) *BQTerrace* (1920×1080)**Fig. 3** SSIM-rate performance curves for low delay configuration (a) *BQSquare* (416 × 240) (b) *BasketballPass* (416 × 240) (c) *PartyScene* (832 × 480) (d) *BQTerrace* (1920 × 1080)

Fig. 4 Visual quality comparisons between HM encoded image and the proposed RDO.

multiplied by the spatial resolutions. The actual SSIM value for each frame is calculated as the average of all SSIM values for the non-overlapped 8×8 blocks that are partitioned from the frame. As shown in Fig. 2-(a), the proposed SSIM model shows high correlation with the actual SSIM with 0.996 for all test sequences. As shown in Fig. 2-(b), for all spatial resolutions, the proposed SSIM model also shows the high correlations with 0.999.

To evaluate the proposed rate distortion optimization coding, our proposed method is integrated into HM16.0 HEVC reference software [9]. Eighteen test sequences from small to high spatial resolution with versatile signal characteristics are used for experiments. Four 416×240 , five 832×480 and five 1920×1080 and two 2560×1600 sequences are tested. Low Delay P and Random Access configurations are used for the experiments. The QP values are 22, 27, 32, and 37. The proposed rate-distortion optimization scheme is compared with Yeo's and Huang's methods in terms of SSIM-rate performance in BDBR [33]. A negative value in BDBR implies that the proposed scheme achieves the bit rate reduction at the same visual quality.

Table 1 shows the rate distortion performances in BDBR where the anchor is HM16.0. As shown in Table 1, the proposed coding scheme achieves significant coding gains against MSE-based RDO in HM16.0 ranging from 0.9% to 15.2% for Low Delay and up to 10.2% for Random Access configuration, respectively and the average coding gain is 6.25%. The results for the Random Access case are

relatively limited compared to the Low Delay P case. It can be inferred that since significantly more number of SKIP blocks does not contribute accurate rate and distortion estimation, the coding gains are reduced. This improvement on coding gains could be obtained because the adaptive Lagrange multiplier is used in the proposed scheme. In addition, the proposed distortion model in (20) reflects the adaptivity of CU depth in each frame. Moreover, the estimation for the parameter of Laplacian PDF and the size of blocks in each CU depth provide accurate rate and SSIM estimations. On the other hands, it can be observed that Yeo's [19] and Huang [13]'s methods failed to produce coding gains in terms of SSIM. Both methods were originally developed for previous coding standard such as H.264/AVC so that they could not efficiently work for more complicated coding structure in HEVC. The Yeo's method is optimized for the identical coding block size as H.264/AVC while HEVC has various types of coding block from 8×8 to 64×64 pixel size. For the integration of the Yeo's method into HEVC, the local variances of the original image are obtained according to CTU. In addition, we found that Huang's method suffers from the error propagation during encoding.

The Rehman's method, which was proposed for HEVC is also compared. The perceptual coding efficiency improvements are not significant compared to the proposed method as shown in Table 1. In particular, coding efficiency improvements for high spatial resolution are relatively limited. Figure 3 illustrates the rate-SSIM curves

for the proposed method, methods from Yeo and Huang and method used in HM16.0 for *BQSquare*, *BasketballPass*, *ParkScene* and *BQTerrace* Sequences, respectively. There are significant improvements based on the proposed scheme for all test sequences as shown in Fig. 3. Especially for *BQSquare* sequences, the proposed schemes shows more than 40% bit rate reduction. As shown in Table 1, it is also observed that the proposed method outperforms the HM16.0 and the other exiting methods. Subjective quality assessment is also investigated. First, the reconstructed images are compared. Figure 4 shows the 34-th and 2-th cropped images of *PartyScene* (832×480) and *BasketballPass* (416×240) sequences for an original and reconstructed images by HM16.0 and the proposed method, respectively.

We encoded the test sequences so that almost identical bits by two coding methods are generated. In Fig. 4-(b), Fig. 4-(c) and 4-(e), Fig. 4-(f), it is observed that the reconstructed frame by the proposed RDO scheme shows more improved visual quality and preserves more details in the edges although the PSNR value is lower than that of HM16.0-encoded frame. As shown in Fig. 4-(b) and -(c), the proposed method shows better subjective quality of the shape around hands. Similarly, better subjective quality of the proposed method than the anchor especially in the head of the player on the right in Fig. 4-(e) and -(f) is observed. As shown in Fig. 4-(b) and -(c), the proposed method shows better subjective quality around hands. Similarly, it can be seen that the proposed method shows better quality especially in the head of the player on the right in Fig. 4-(e) and -(f). It implies that our proposed RDO scheme performs the

encoding process of HEVC in perceptual distortion and rate optimization sense.

We also performed the subjective quality evaluation based on a two-alternative forced choice (2AFC) [33], [34] which is known as the most common psychophysical test. In the experiment, a subject is given a pair of video sequences encoded by the original HM16.0 and the proposed method. Then, he/she is asked to select which has better visual quality. Total 12 non-expertise subjects in this field were participated in the subjective quality evaluation. We choose six pairs of sequences which have same levels of SSIM index. Each pair is repeated six times in random orders. Consequently, we obtained 36 2AFC readings for each result and subject. The viewing distance is 160cm for HD sequences

Table 2 Rates and SSIM of sequences used in subjective quality test (BQ: *BQSquare*, BP: *BasketballPass*, PS: *PartyScene*, BD: *BasketballDrive*, BT: *BQTerrace*, PSC: *ParkScene*)

Seq.	Set 0				Set 1			
	Proposed		HM16.0		Proposed		HM16.0	
	SSIM	Rates	SSIM	Rates	SSIM	Rates	SSIM	Rates
1 BQ	0.871	440.0	0.870	461.2	0.911	960.1	0.910	1250.1
2 BP	0.850	221.1	0.849	231.2	0.916	490.2	0.916	518.2
3 PS	0.838	1610.6	0.839	1790.5	0.913	3805.6	0.914	4214.5
4 BD	0.853	1650.1	0.853	1757.1	0.912	2562.1	0.912	3242.4
5 BT	0.864	1621.9	0.863	1800.1	0.912	6301.1	0.912	6423.1
6 PSC	0.855	980.2	0.855	1159.6	0.921	2296.1	0.920	2489.1

(a) Ordered by sequence for Set 0

(b) Ordered by subjects for Set 0

(c) Ordered by sequence for Set 1

(d) Ordered by subjects for Set 1

Fig. 5 Error plot bar for subjective test. (a) Ordered by sequence for Set 0. (b) Ordered by subjects for Set 0. (c) Ordered by sequence for Set 1. (d) Ordered by subjects for Set 1

Table 3 Encoding time comparison between original HEVC and the proposed method

Sequences		ΔT (%)
1	<i>BQSquare</i>	5.7
2	<i>BasketballPass</i>	5.9
3	<i>PartyScene</i>	6.2
4	<i>BasketballDrill</i>	7.0
5	<i>BQTerrace</i>	7.3
6	<i>ParkScene</i>	8.5
Average		7.1

and 80 cm for non-HD sequences. We set up two types of experimental sets. The video coded at the lowest bitrate range are test using QP 37 for Set 0 while the video coded at the second lowest bitrate range using QP 32 are test for Set 1. Table 2 shows the SSIM and rate values for each set. As shown in Table 2, the sequences with identical SSIM but different rate values are given to select for each experiment Set. The six sequences from various spatial resolutions are selected for subjective evaluation for the experiments, other sequences shows similar tendency in the results. As observed in Fig. 5, the average percentage that the subjects have selected the proposed method is around 50% for Set 0 and Set 1 where they are 51.8% and 48.8% in subjects and 50.5% and 50.8% in sequences, for Set 0 and Set 1 conditions, respectively. This implies that the subjective quality differences between two methods are not significantly different. However, the rates produced by the proposed method are less than those by the original HM16.0 from 0.2% to -15.2%. The proposed method achieves identical subjective quality with significantly reduced bit amounts.

For coding complexity evaluation, encoding times are measure as

$$\Delta T(\%) = \frac{T_{\text{Pro}} - T_{\text{Org}}}{T_{\text{Org}}} \times 100(\%) \quad (37)$$

where T_{Org} and T_{Pro} indicate the total encoding time for conventional MSE based rate-distortion optimization of HEVC and the proposed method, respectively. A computer with Intel core i-7 4820 3.7GHz and RAM 8GB was used for simulation. Table 3 shows encoding time increases of the proposed RDO scheme. As shown in the table, encoding times slightly increase from about 5% to about 8%. It is due to the fact that the proposed method spends overhead complexity to compute block SSIM values as well as statistical values such as block variances for λ_{ssim} calculation. While most of complexity overhead comes from computation of block SSIM values, complexity for processing time and data collection for new Lagrange calculation are negligible.

Figure 6 shows the comparisons for proportion of selected CU sizes between original and the proposed method for different QP values. The data is obtain from the average values of the encoding results of *BasketballDrive* (1920×1080), *BQTerrace* (1920×1080), *PeopleOnStreet* (2560×1600) sequences when QP22 and QP32 are used.

Fig. 6 Comparisons for proportion of selected CU sizes between HM16.0 and the proposed method for different QP values.**Table 4** BDBR (%) results when Eq. (35) is used or not used.

Sequences	Not using (35) (%)	Using (35) (%)
<i>BQSquare</i>	-9.8	-11.3
<i>BlowingBubbles</i>	-1.3	-4.6
<i>BasketballDrill</i>	-10.1	-11.2
<i>PartyScene</i>	-7.7	-10.3
<i>Vidyo1</i>	-0.8	-0.9
<i>Vidyo3</i>	-1.6	-1.6
<i>ParkScene</i>	-2.1	-4.8
<i>BasketballDrive</i>	-13.4	-15.2
<i>Cactus</i>	-8.4	-9.2
Average	-6.1	-7.8

As shown in the figure, the selected partitions of large CU sizes such as 64×64 and 32×32 are increased by 13% and 8% for QP22 and QP 32, respectively. This observation supports that the proposed method leads significant bit savings under the same level of SSIM values by selecting relatively larger block sizes. This result is consistent with results in Table 2, where bit savings via the proposed method can be obtained under the same SSIM values.

In the proposed method, we introduce a scheme to avoid error propagation for more reliable perceptual rate-distortion optimization in Sect. 2.2. In order to investigate the effect the scheme, an experiment is performed with or without (35), which is a Lagrange multiplier clipping method according to the ratio of two variances (R_n) for two successive frame difference. Table 4 shows the BDBR (%)

results when Eq. (35) is used or not used. The test sequences with high temporal motions such as *ParkScene* and *PartyScene* as well as ones with relatively static motions such as *Vidyo1* are compared. As shown in Table 4, the BDBRs for the *PartyScene* and *ParkScene* sequence with high motions are reduced with 2.6% and 2.7% points, respectively when Eq. (35) is turned off while the BDBR values for *Vidyo1* and *Vidyo3* sequences are rarely changed although Eq. (35) is not used. It implies that the scheme to avoid error propagation using Eq. (35) effectively works when the statistical model based on the Laplacian distribution might not be valid by high motions or abrupt changes.

4. Conclusions

In this paper, a perceptual distortion based rate distortion optimized coding scheme is proposed with a new SSIM model for HEVC. For this, an SSIM-quantization model is derived under the high resolution quantization assumption by taking into account the different statistical characteristics for different CU depth levels of HEVC. It is shown in the experiments that the proposed SSIM-quantization model produces highly correlated SSIM values with the actual SSIM values. This results from accurately derived Lagrange multiplier in perception distortion and rate optimization based HEVC encoding. Our rate model is derived based on the entropy, which is computed from Laplacian PDF of transform coefficients of predicted residues. Based on the proposed SSIM and rate models, a Lagrange multiplier for rate distortion optimization coding can be obtained. The experimental results show that the proposed perceptual distortion based rate distortion optimization coding scheme achieves the average rate reduction with about 6.0% compared to the MSE-based existing method. In addition, the proposed scheme shows identical subjective quality even with significantly reduced bit amounts.

Acknowledgements

This study was supported by research fund from Chosun University, 2016.

References

- [1] 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, April 2004.
- [2] T. Wiegand, B. Bross, W.-J. Han, J.-R. Ohm, and G.J. Sullivan, "Working draft 3 of high efficiency video coding," JCTVC-E603, JCTVC of ISO/IEC and ITU-T, Geneva, Swiss, March 2011.
- [3] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol.13, no.7, pp.560–576, July 2003.
- [4] G.J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol.22, no.12, pp.1649–1668, Dec. 2012.
- [5] I.-K. Kim, J. Min, T. Lee, W.-J. Han, and J.-H. Park, "Block partitioning structure in the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol.22, no.12, pp.1697–1706, Dec. 2012.
- [6] G.J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol.15, no.6, pp.74–90, 1998.
- [7] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," *IEEE Int. Conf. Image Process. (ICIP)*, Thessaloniki, Greece, vol.3, pp.542–545, 2001.
- [8] JVT. H.264/AVC Reference Software (JM13.1), Dec. 2007 [Online]. Available: <http://iphome.hhi.de/suehring/tm1>.
- [9] HM reference software 16.0, [Online] Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-11.0.
- [10] X. Li, N. Oertel, A. Hutter, and A. Kaup, "Laplace distribution based Lagrangian rate distortion optimization for hybrid video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol.23, no.7, pp.193–205, Feb. 2009.
- [11] L. Chen and I. Garbacea, "Adaptive Lambda estimation in Lagrangian rate-distortion optimization for video coding," *Proc. Visual Commun. Image Process. (VCIP)*, San Jose, CA, Jan. 2006.
- [12] C. Pang, O.C. Au, J. Dai, and F. Zou, "Frame Complexity Guided Lagrange Multiplier Selection for H.264 Intra-Frame Coding," *IEEE Signal Process. Lett.*, vol.18, no.12, pp.733–736, Dec. 2011.
- [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 Trans. Circuits Syst. Video Technol.*, vol.20, no.11, pp.1614–1624, Nov. 2010.
- [14] 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.
- [15] 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, April 2012.
- [16] 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, April 2013.
- [17] A. Rehman and Z. Wang, "SSIM-inspired perceptual video coding for HEVC," *proc. Int. Conf. Multimedia Expo (ICME) 2012*, pp.497–502, July 2012.
- [18] C. Yeo, H.L. Tan, and Y.H. Tan, "On rate distortion optimization using SSIM," *IEEE Trans. Circuits Syst. Video Technol.*, vol.23, no.7, pp.1170–1181, July 2013.
- [19] C. Yeo, H.L. Tan, and Y.H. Tan, "On rate distortion optimization using SSIM," *proc. Int. Conf. Acoustics, Speech Signal Process. (ICASSP) 2012*, pp.833–836, March 2012.
- [20] S. Wang, A. Rehman, K. Zeng, J. Wang, and Z. Wang, "SSIM-motivated two-pass VBR coding for HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol.27, no.10, pp.2189–2203, Oct. 2017.
- [21] W. Gao, S. Kwong, Y. Zhou, and H. Yuan, "SSIM-Based Game Theory Approach for Rate-Distortion Optimized Intra Frame CTU-Level Bit Allocation," *IEEE Trans. Multimedia*, vol.18, no.6, pp.988–999, June 2016.
- [22] J. Kim, S.-H. Bae, and M. Kim, "An HEVC-compliant perceptual video coding scheme based on JND models for variable block-sized transform kernels," *IEEE Trans. Circuits Syst. Video Technol.*, vol.25, no.11, pp.1786–1800, Nov. 2015.
- [23] B. Lee and M. Kim, "Modeling Rates and Distortions Based on a Mixture of Laplacian Distributions for Inter-Predicted Residues in Quadtree Coding of HEVC," *IEEE Signal Process. Lett.*, vol.18, no.10, pp.571–574, Oct. 2011.
- [24] B. Lee and M. Kim, "No-reference PSNR estimation for HEVC encoded video," *IEEE Trans. Broadcast.*, vol.59, no.1, pp.20–27, Jan. 2013.
- [25] B. Lee, M. Kim, and T.Q. Nguyen, "A frame-level rate control scheme based on texture and non-texture rate models for High Efficiency Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol.24, no.3, pp.465–479, March 2014.
- [26] B. Lee, J. Jung, and M. Kim, "An all-zero block detection scheme for low-complexity HEVC encoders," *IEEE Trans. Multimedia*, vol.17,

- no.7, pp.1257–1268, July 2016.
- [27] B. Lee, S. Ahn, and M. Kim, “A novel distortion model for quadtree coding in high efficiency video coding,” *Proc. SPIE-IS&T Electronic Imaging 2012*, vol.8305, pp.830503-1–830503-9, 2012.
 - [28] G.J. Sullivan and S. Sun, “On dead zone plus uniform threshold scalar quantization,” *Proc. Visual Commun. Image Process. (VCIP)*, July 2005.
 - [29] S.S. Channappayya, A.C. Bovik, and R.W. Heath Jr., “Rate Bounds on SSIM Index of Quantized Images,” *IEEE Trans. Image Process.*, vol.17, no.9, pp.1624–1639, Sept. 2008.
 - [30] W. Gao, S. Kwong, H. Yuan, and X. Wang, “DCT Coefficient Distribution Modeling and Quality Dependency Analysis Based Frame-Level Bit Allocation for HEVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol.26, no.1, pp.139–153, Jan. 2016.
 - [31] H. Yuan, S. Kwong, X. Wang, W. Gao, and Y. Zhang, “Rate Distortion Optimized Inter-View Frame Level Bit Allocation Method for MV-HEVC,” *IEEE Trans. Multimedia*, vol.17, no.12, pp.2134–2146, Dec. 2015.
 - [32] G. Bjøntegaard, “Calculation of average PSNR differences between RD-Curves,” ITU-T SG16 Q.6 Document, VCEG-M33, April 2001.
 - [33] S. Winkler, “Analysis of public image and video databases for quality assessment,” *IEEE J. Sel. Topics Signal Process.*, vol.6, no.6, pp.616–625, Oct. 2012.
 - [34] Recommendation 500-10: Methodology for the subjective assessment of the quality of television pictures, ITU-R Standard Rec. BT.500, 2012.

Bumshik Lee received the B.S. degree in Electrical Engineering from Korea University, Seoul, Korea, in 2000, and the M.S. and Ph.D. degrees in Information and Communications Engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2006 and 2012, respectively. He was a Research Professor in KAIST, Korea in 2014 and a Postdoctoral Scholar in University of California, San Diego (UCSD), CA, USA from 2012 to 2013. He was a Principal Engineer in Advanced Standard R&D Lab., LG Electronics, Seoul from 2015 to 2016. In 2016, he joined as an Assistant Professor in Department of Information and Communications Engineering, Chosun University, Korea.

His research interests include video compression, video processing, video security and medical image processing etc.

Jae Young Choi received the B.S. degree from Kwangwoon University, Seoul, South Korea, in 2004, and the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2008 and 2011, respectively. In 2008, he was a visiting scholar at the University of Toronto and from 2011 to 2012, he was a Post-Doctoral Researcher with the University of Toronto. He was a Postdoc Fellow with the University of Pennsylvania (UPenn) from 2012 to

2013. He was a Senior Engineer with SAMSUNG ELECTRONICS from 2013 to 2014. He is currently an Assistant Professor with Division of Computer & Electronics Systems Engineering, Hankuk University of Foreign Studies. His research interests include pattern recognition, machine learning, image processing, and computer vision. Especially, he has developed several pioneering algorithms for automatic face recognition using facial color information. He is the author or co-author of over 80 refereed research publications in the aforementioned research areas. Dr. Choi was the recipient of the Samsung HumanTech Thesis Prize in 2010. He is a member of the IEEE. He served as a regular reviewer in many journals including *IEEE Trans. Image Processing*, *IEEE Trans. Multimedia*, *IEEE Trans. System, Man and Cybernetics*.