SSIM-Based Coarse-Grain Scalable Video Coding

Tiesong Zhao, *Member, IEEE*, Jiheng Wang, *Student Member, IEEE*, Zhou Wang, *Fellow, IEEE*, and Chang Wen Chen, *Fellow, IEEE*

***Abstract*—We propose an improved coarse-grain scalable video coding (SVC) approach based on the structural similarity (SSIM) index as the visual quality criterion, aiming at maximizing the overall coding performance constrained by user-defined quality weightings for all scalable layers. First, we develop an interlayer rate-SSIM dependency model, by investigating bit rate and SSIM relationships between different layers. Second, a reduced-reference SSIM-Q model and a Laplacian R-Q model are introduced for SVC, by incorporating the characteristics of hierarchical prediction structure in each layer. Third, based on the user-defined weightings and the proposed models, we design a rate-distortion optimization approach to adaptively adjust Lagrange multipliers for all layers to maximize the overall rate-SSIM performance of the scalable encoder. Experiments with multiple layers, different layer weightings, and various videos demonstrate that the proposed framework can achieve better rate-SSIM performance than single layer optimization method, and provide better coding efficiency as compared to the conventional SVC scheme. Subjective tests further demonstrate the benefits of the proposed scheme.**

***Index Terms*—Scalable video coding (SVC), coarse-grain scalability (CGS), structural similarity (SSIM), rate-distortion optimization (RDO), Lagrange multiplier (LM).**

I. Introduction

In THE past decades, digital video coding technologies have been greatly improved, represented by state-of-the-art standards, which include H.264 Advanced Video Coding (H.264/AVC) [1], High Efficiency Video Coding (HEVC) [2], Scalable Video Coding (SVC) [3] and Multiview Video Coding (MVC) [4]. Many techniques contribute to the improvement in performance, including variable partition sizes [5], motion search with multiple reference frames [6], entropy coding [7], [8], deblocking filter [9], rate control [10], coding tree unit [11], and coding unit merging [12]. These techniques are incorporated into a sophisticated video coding scheme, which can be characterized by a hybrid model of motion-handling and picture-coding, with a Rate-Distortion Optimization (RDO) process to minimize the distortion subject to a constraint on Bit Rate (BR) [13]. In practice, Lagrange optimization is applied to combine the distortion (D) and BR to a Rate-Distortion (RD) cost, with a Lagrange Multiplier (LM). The objective of RDO is hence to minimize the RD cost using the afore-mentioned techniques.

In a hybrid encoder, the perceived distortion and its inverse, visual quality, are difficult to measure due to the complexity of the human perceptual system and the statistics of natural videos. In state-of-the-art video coding standards [1]–[4], Mean Squared Error (MSE) and Peak-Signal-to-Noise Ratio (PSNR) are commonly adopted as distortion and quality measures, respectively, which have resulted in many RDO algorithms, including optimal bit allocation with multiple partition sizes [14], dependent joint RDO using soft decision quantization [15], RDO based on Laplacian coding residuals [16], and RD optimized transform [17]. Nevertheless, these existing measures have been widely criticized for their low correlations with perceived quality [18]. On the other hand, although some state-of-the-art video quality measures, such as Video Quality Model (VQM) [19] and MOtion-based Video Integrity Evaluation (MOVIE) index [20], can achieve relatively good performance, the computational complexities are extremely high, making them difficult to be incorporated in the design of video encoders [21].

In recent years, the structural similarity (SSIM) index [22], [23] has been increasingly popular as a replacement of MSE/PSNR for the evaluation and optimization of video codecs. In addition to its competitive quality prediction performance and low complexity [18], [21], [24], it also has a number of additional desirable properties. It is differentiable, locally convex, quasi convex, and its direct variations are shown to be valid distance metrics [25]. It can produce a quality map that indicates local quality variations, providing a useful guiding tool in bit allocation of video coding schemes. It also saturates at high rate, which is consistent with the behavior of the visual system [26].

Recently, SSIM has been incorporated into Motion Estimation (ME), mode selection and rate control algorithms, to enhance compression efficiency whilst maintaining the perceptual quality [27]–[38]. Mai *et al.* [27], [28] and Aswathappa and Rao [29] independently proposed SSIM-based intra prediction mode decision methods, with a fixed LM for each Quantization Parameter (Qp). In [30], SSIM was employed to improve ME process in H.264 to achieve more BR reduction with the same perceptual video quality. Based on SSIM and the derivation of LM, Yang *et al.* [31], [32] developed algorithms to improve inter prediction and RDO processes, respectively. Ou *et al.* [33] proposed an optimal

Manuscript received May 4, 2014; revised March 24, 2015; accepted April 1, 2015. Date of publication May 8, 2015; date of current version June 4, 2015.

T. Zhao, J. Wang, and Z. Wang are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: ztiesong@uwaterloo.ca).

C. W. Chen is with the Department of Computer Science and Engineering, University at Buffalo, State University of New York, Buffalo, NY 14260 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2015.2424012

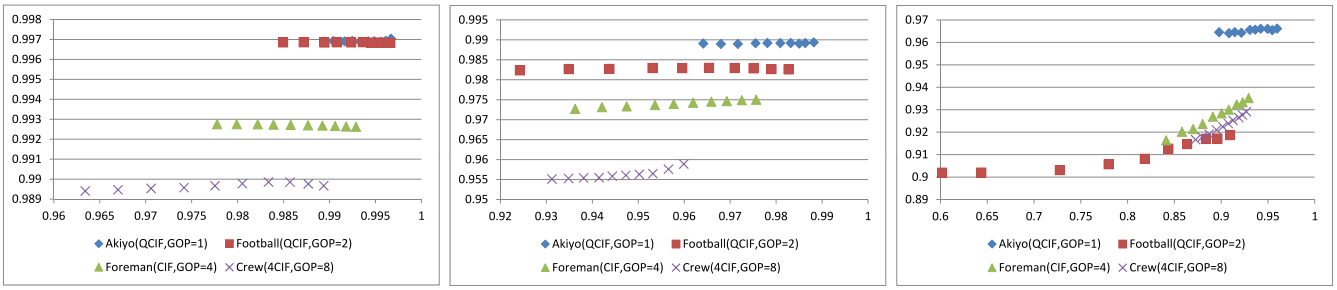


Fig. 1. SSIM dependency between BL and EL in a two-layer case. Horizontal axis: SSIM of BL. Vertical axis: SSIM of EL. (a) $Qp_E = 10$. (b) $Qp_E = 20$. (c) $Qp_E = 30$.

SSIM-based bit allocation and rate control scheme for H.264. To achieve better coding performance, the LM is further determined by content-adaptive parameters [34]–[38]. Huang *et al.* [34], [36] and Chen *et al.* [35] developed a perceptual LM estimation method, in which the RD point of previous coded frame is horizontally and vertically projected to a pre-modeled RD curve, and then the LM is determined by the slope between the two projection points. Wang *et al.* [37], [38] proposed a Reduced-Reference (RR) statistical SSIM estimation method, which is further utilized to develop an SSIM-Q model. Based on this SSIM-Q model and a Laplacian R-Q model [16], the perceptual LM is adaptively determined by video content. Besides, at MacroBlock (MB) level, the LM is further adjusted based on an information theoretical approach. In [39], an MB level perceptual mode selection scheme and a frame level global quantization matrix optimization method are developed based on a divisive normalization scheme.

In this paper, we propose SSIM-based Coarse-Grain Scalability (CGS) coding method to improve the Rate-SSIM (R-S) performance of the scalable video codec SVC [3] by investigating the inter-layer R-S dependency. SVC can produce scalable bit streams with only one encoder and adapt to various device capabilities, network conditions and client applications. There are three types of scalability in SVC, namely temporal, spatial and quality scalability, respectively, where quality scalability supports two modes, known as the CGS and the Medium-Grain Scalability (MGS) [40]. The proposed approach contributes to SVC in the following three aspects. First, we incorporate SSIM as the distortion measure in the current CGS encoder, and develop an inter-layer R-S dependency model to characterize the relationships between BRs and SSIMs of different CGS layers. Second, we introduce SSIM-Q and R-Q models to SVC, to determine adaptive LMs with parameter predictions for all temporal and quality layers. Third, by combining SSIM-Q, R-Q and inter-layer R-S dependency models, we propose an SSIM-based RDO method for CGS encoding, which can maximize the overall R-S performance with user-defined weights for different CGS layers.

The remaining of the paper is organized as follows. In Section II, the inter-layer R-S dependency model among different CGS layers is presented. In Section III, the overall RDO scheme with user-defined weights for different layers is described. Various sequences with multiple CGS layers, different layer weights and Qps are tested and discussed, and subjective test results are demonstrated in Section IV. Finally, Section V concludes the paper.

II. Interlayer R-S Dependency in CGS Encoder

In SVC, the concept of "layer" is introduced to represent different sub-streams. A Base Layer (BL) can be decoded independently, but with lower reconstruction quality comparing with the complete bit stream. An Enhancement Layer (EL) can only be decoded by incorporating information from BL and lower ELs, but can achieve reconstruction quality better than that in lower layers. In temporal and spatial scalability, a layer represents the source content with a reduced frame rate and picture size, respectively; while in quality scalability, all layers are with identical frame rate and picture size, but at different quality levels. The intra-layer dependency between frames, also known as temporal coding dependency, has been widely studied for various coding structures of H.264/AVC [41]–[43]. In this work, we investigate the inter-layer R-S dependency first to improve joint-layer optimization of SSIM-based CGS coding.

A. R-S Dependency Between BL and EL

The local SSIM index of two local image patches x and y is defined as [23]

$$\mathrm{SSIM}(\mathrm{x},\mathrm{y}) = \frac{(2\mu_\mathrm{x}\mu_\mathrm{y} + \mathrm{C}_1)(2\sigma_\mathrm{xy} + \mathrm{C}_2)}{(\mu_\mathrm{x}^2 + \mu_\mathrm{y}^2 + \mathrm{C}_1)(\sigma_\mathrm{x}^2 + \sigma_\mathrm{y}^2 + \mathrm{C}_2)}, \quad (1)$$

where μ_x, μ_y are the mean values of the two patches, σ_x, σ_y and σ_xy are the standard derivations of x, y, and the cross correlation between the two patches, respectively; C_1 and C_2 are positive constants. The frame level and sequence level SSIMs are computed by averaging the SSIM values of all local patches and all frames, respectively.

To investigate the inter-layer R-S dependency, we first implement the SSIM-based H.264/AVC RDO scheme [38] in SVC with CGS. We change Qps of both EL and BL to observe how SSIM and BR change at different layers. In the three subfigures of Fig. 1, we set Qp of EL, Qp_E, as 10, 20 and 30, respectively; in each subfigure, we gradually increase the Qp of BL from $Qp_E + 1$ to $Qp_E + 10$, which is sufficiently large to account for the working range of RDO algorithms.

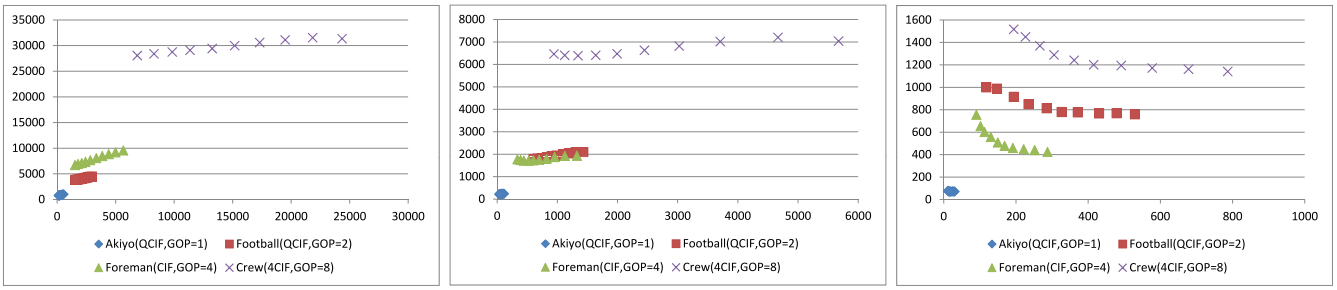


Fig. 2. BR dependency between BL and EL in a two-layer case. Horizontal axis: BR of BL. Vertical axis: BR of EL. (a) $Qp_E = 10$. (b) $Qp_E = 20$. (c) $Qp_E = 30$.

Four sequences (*Akiyo*, *Foreman*, *Football* and *Crew*) are tested to show the SSIM relationship between BL and EL; three resolutions are included, which are Common Intermediate Format (CIF), Quarter CIF (QCIF) and 4× CIF (4CIF); a wide range of Group-Of-Pictures (GOP) sizes is examined from 1 to 8, with 33 frames coded. From each subfigure, we observe that there exists an approximately linear relationship between the SSIM of EL, SSIM_E, and the SSIM of BL, SSIM_B, such that

$$\frac{\Delta\text{SSIM}_E}{\Delta\text{SSIM}_B} \approx \mathscr{S}(Qp_E, Qp_B), \tag{2}$$

where Δ denotes small variations, and the slop $\mathscr{S}(\cdot)$ is approximately a constant for a specific Qp_E and a large range of Qp_B, as shown in Fig. 1. Similar conclusions can also be drawn on the other benchmark sequences, GOP sizes and Qp settings. It is also observed that with a fixed Qp_E in Fig. 1, $\mathscr{S}(\cdot)$ is almost independent from Qp_B. Thus, in this work, we set the slop $\mathscr{S}(\cdot)$ to be a function of Qp_E only, such that $\mathscr{S}(Qp_E, Qp_B) = \mathscr{S}(Qp_E) = f(Q_E)$, where Q_E represents the corresponding Q_{step} of Qp_E.

As Q_E increases from Fig. 1(a)–(c), the slope also increases, which inspires us to set it as an increasing function of Q_E. We use a simple empirical approximation of $f(Q_E)$ given by

$$f(Q_E) \approx a \cdot Q_E, \tag{3}$$

where a is a positive real number. In the extreme case, when $Q_E = 0$, $f(Q_E)$ is set to be zero, which corresponds to $\text{SSIM}_E = 1$ and $\Delta\text{SSIM}_E \approx 0$ when no quantization is applied.

The BR dependency between BL and EL is observed in Fig. 2. Similarly, in a small neighborhood, there exists approximately a linear relationship between the changes of BRs in BL and EL, that is,

$$\frac{\Delta R_E}{\Delta R_B} \approx g(Q_E). \tag{4}$$

It can be observed from Fig. 2 that $g(Q_E)$ is a decreasing function of Q_E, and is positive with a small Q_E and negative when Q_E is large. Empirically, we approximate $g(Q_E)$ by

$$g(Q_E) \approx b - c \cdot Q_E, \tag{5}$$

where b and c are positive real numbers.

The inner mechanism of the above models may be explained with the inter-layer predictions including *prediction of MB modes and associated motion parameters* and *prediction of the residual signal* [40]. Due to error propagation, a lower SSIM of BL will definitely result in a lower SSIM of EL, when inter-layer predictions are employed. Therefore, the SSIM of EL is positively correlated to that of BL in Fig. 1, and it leads to a positive a in Eq. (3). When BR of BL varies, the increment of BR of EL may be positive or negative, depending on the trade-off of the RDO process. As we observed in Fig. 2, there exists a positive correlation between BRs of BL and EL with small distortion (i.e., high SSIM), and a negative correlation with large distortion (i.e., low SSIM).

B. R-S Dependency Between Neighboring ELs

To develop RDO scheme for multiple CGS layers, the R-S dependency between one BL and more than one EL is also studied. We test three CGS layers with one BL and two ELs, denoted by lower EL (E1) and higher EL (E2), respectively, where E2 can be predicted either from BL or E1 [40]. The former case has been discussed in Section II-A. In this work we focus on the latter case, which is also the default and frequently-used setting in SVC reference software JSVM. In this case, the inter-layer prediction of E2 is from E1 only, hence there is no direct R-S dependency between E2 and BL. Since BL-E1 dependency has been studied in Section II-A, we exploit E1-E2 dependency in this section and thus BL-E2 dependency can be derived with BL-E1 and E1-E2 dependencies.

The E1-E2 dependency, also called R-S dependency between neighboring ELs, is studied with the same coding configurations to those of Figs. 1 and 2. We change the Qps of E2 and E1 to observe the SSIM and BR changes. The SSIM and BR dependencies between E1 and E2 are shown in Figs. 3 and 4, respectively. From the two figures, we conclude that

$$\frac{\Delta\text{SSIM}_{E2}}{\Delta\text{SSIM}_{E1}} \approx 0, \tag{6}$$

and

$$\frac{\Delta R_{E2}}{\Delta R_{E1}} \approx d, \tag{7}$$

where d is a positive real number between 0 and 1. We use Eq. (6) mainly for two reasons. Firstly, the SSIM_{E2} change is

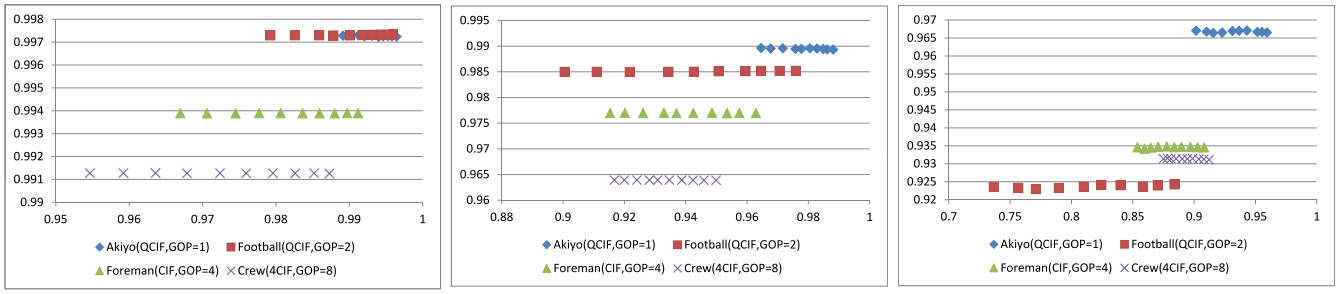


Fig. 3. SSIM dependency between neighboring ELs in a three-layer case. Horizontal axis: SSIM of $E1$. Vertical axis: SSIM of $E2$. (a) Qp of $E2$ equals 10. (b) Qp of $E2$ equals 20. (c) Qp of $E2$ equals 30.

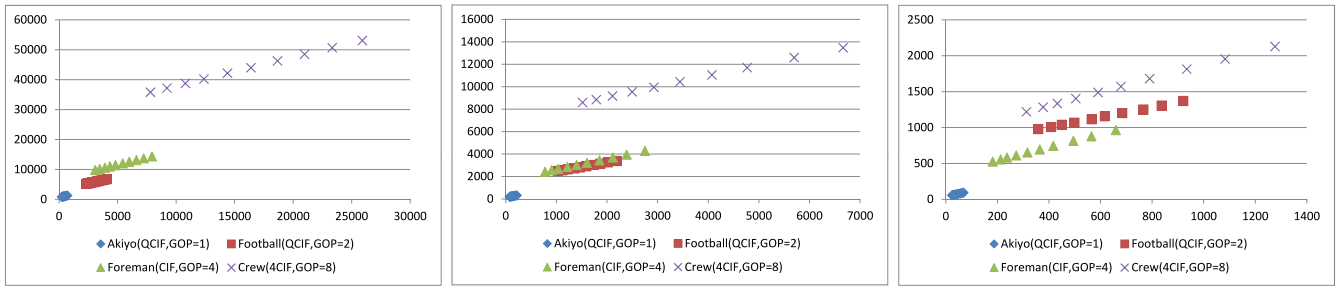


Fig. 4. BR dependency between neighboring ELs in a three-layer case. Horizontal axis: BR of $E1$. Vertical axis: BR of $E2$. (a) Qp of $E2$ equals 10. (b) Qp of $E2$ equals 20. (c) Qp of $E2$ equals 30.

relatively small when SSIM_{E1} changes, as shown in Fig. 3, which allows us to reduce the complexity by discounting its impact; secondly, due to the saturation property of SSIM index, the SSIM in higher CGS layer (which is usually close to 1) is more stable even if the SSIM of a lower layer changes.

The R-S dependency between neighboring ELs can be approximately considered as a special case of R-S dependency between BL and EL, where the distortions are very small at higher CSG layers. In such a case, Eqs. (3) and (5) can be approximated by 0 and a constant, respectively.

Thereafter, for CGS layer $L > 0$, the inter-layer R-S dependency model is defined as

$$\frac{\Delta \text{SSIM}_L}{\Delta \text{SSIM}_{L-1}} = \partial_D(L) = \begin{cases} a \cdot Q_L & \text{if } L = 1, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

$$\frac{\Delta R_L}{\Delta R_{L-1}} = \partial R(L) = \begin{cases} b - c \cdot Q_L & \text{if } L = 1, \\ d & \text{otherwise,} \end{cases} \tag{9}$$

where Q_L, SSIM_L, and R_L denote Q_{step}, SSIM and BR of layer L, respectively; δ_D and δ_R are dependency functions. The model parameters a, b, c and d may be different for different sequences and GOP sizes. In our work, these parameters are adjusted based on a set of 4CIF sequences with Qps from 10 to 30, and are finally set to be $a = 0.001$, $b = 0.2$, $c = 0.002$ and $d = 0.5$, respectively, which also results in good performances for CIF and High Definition (HD) sequences, as will be shown in Section IV.

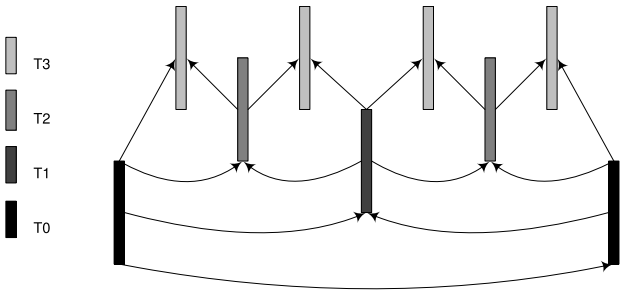


Fig. 5. Illustration of an HBP structure with a GOP of size 8.

III. SSIM-Based RDO for CGS Encoding

Based on the inter-layer R-S dependency model introduced above, here we propose an RDO method for CGS encoder, to maximize the weighted sum of R-S performances of all CGS layers. In each layer of CGS, SSIM-Q and R-Q models are utilized for all hierarchical levels, in a Hierarchical-B-Pictures (HBP) structure [44]. An example of HBP is shown in Fig. 5, where pictures with higher T values are predicted from those with the same or lower T values, which also forms temporal scalability, because lower T-valued pictures could be decoded independently with lower frame rates.

A. SSIM-Based RDO With User-Defined Weights

In [13], the RDO problem in video coding is formalized as

$$\min\{D\}, s.t. R < R_c, \tag{10}$$

where D and R denote the distortion and BR, respectively; R_c is the constraint on the number of bits used. This problem can be addressed with Lagrange optimization,

$$\min\{J\}, \text{ where } J = D + \lambda R. \tag{11}$$

The Lagrange function J is also called RD cost. In an SSIM-based RDO process, the distortion is defined as the opposite of visual quality,

$$D = 1 - \text{SSIM}. \tag{12}$$

Based on Eqs. (11) and (12), the SSIM-based R-S cost is defined as

$$J = (1 - \text{SSIM}) + \lambda R. \tag{13}$$

Ulteriorly, for a block, the R-S cost is

$$J_{blk} = \sum_{k \in blk} (1 - \text{SSIM}_k) + \lambda R_{blk}, \tag{14}$$

where SSIM_k and R_{blk} denote the SSIM of pixel k and the total bits of the block, respectively.

For CGS coding with multiple layers $L = 0, 1, \ldots, L_{max}$, let the user-defined distortion and bit weights for layer L be w_L^D and w_L^R, respectively; $w_L^D > 0$, $w_L^R > 0$, and $\sum_{L=0}^{L_{max}} w_L^D = \sum_{L=0}^{L_{max}} w_L^R = 1$, then the objective of SSIM-based RDO with multiple CGS layers is

$$\min\left\{\sum_{L=0}^{L_{max}} w_L^D D_L\right\}, s.t. \sum_{L=0}^{L_{max}} w_L^R R_L < R_c, \tag{15}$$

where D_L and R_L denote the distortion of total bits of layer L, respectively; $D_L = 1 - \text{SSIM}_L$. Hence, for CGS layer L, the LM can be determined by

$$\begin{aligned}\lambda_L &= -\frac{\sum_{i=L}^{L_{max}} \left\{ \frac{w_i^D}{\sum_{j=L}^{L_{max}} w_j^D} \cdot \partial D_i \right\}}{\sum_{i=L}^{L_{max}} \left\{ \frac{w_i^R}{\sum_{j=L}^{L_{max}} w_j^R} \cdot \partial R_i \right\}} \\ &= \frac{\sum_{i=L}^{L_{max}} w_i^R}{\sum_{i=L}^{L_{max}} w_i^D} \cdot \frac{\sum_{i=L}^{L_{max}} w_i^D \cdot \partial \text{SSIM}_i}{\sum_{i=L}^{L_{max}} w_i^R \cdot \partial R_i}.\end{aligned} \tag{16}$$

Based on inter-layer SSIM dependency model in Eq. (8), we can obtain that

$$\frac{\partial \text{SSIM}_i}{\partial \text{SSIM}_L} = \prod_{j=L+1}^{i} \delta_D(j), i > L; \tag{17}$$

and based on inter-layer BR dependency model in Eq. (9),

$$\frac{\partial R_i}{\partial R_L} = \prod_{j=L+1}^{i} \delta_R(j), i > L. \tag{18}$$

Substitute Eqs. (17) and (18) into Eq. (16), the LM in layer L can be deduced as

$$\lambda_L = \Theta_L \cdot \frac{\partial \text{SSIM}_L}{\partial R_L} = \Theta_L \cdot \frac{\frac{\partial \text{SSIM}_L}{\partial Q_L}}{\frac{\partial R_L}{\partial Q_L}}, \tag{19}$$

where

$$\Theta_L = \frac{\sum_{i=L}^{L_{max}} w_i^R}{\sum_{i=L}^{L_{max}} w_i^D} \cdot \frac{w_L^D + \sum_{i=L+1}^{L_{max}} \left\{ w_i^D \prod_{j=L+1}^{i} \delta_D(j) \right\}}{w_L^R + \sum_{i=L+1}^{L_{max}} \left\{ w_i^R \prod_{j=L+1}^{i} \delta_R(j) \right\}}. \tag{20}$$

In this work, we define the overall R-S performance as a weighted sum of R-S performances of all CGS layers. In such a case, $w_L^D = w_L^R = w_L$ represents the user-defined R-S weight of layer L, and the total R-S cost can be represented as a weighted sum of R-S costs of all layers,

$$J_{tot} = \sum_{L=0}^{L_{max}} w_L J_L, \tag{21}$$

where J_L is the R-S cost layer L. Substitute $w_L^D = w_L^R = w_L$ into Eq. (20), we have

$$\Theta_L = \frac{1 + \sum_{i=L+1}^{L_{max}} \left\{ \frac{w_i}{w_L} \prod_{j=L+1}^{i} \delta_D(j) \right\}}{1 + \sum_{i=L+1}^{L_{max}} \left\{ \frac{w_i}{w_L} \prod_{j=L+1}^{i} \delta_R(j) \right\}}. \tag{22}$$

B. SSIM-Q and R-Q Models

To determine $\frac{\partial \text{SSIM}_L}{\partial Q_L}$ and $\frac{\partial R_L}{\partial Q_L}$ in Eq. (19), SSIM-Q and R-Q models are employed in each CGS layer. In our work, we use the RR SSIM-Q model [38], which was derived with a DCT domain SSIM index presented by Channappayya *et al.* [45]:

$$\begin{aligned}\text{SSIM(x, y)} = &\left\{ 1 - \frac{(X(0) - Y(0))^2}{X(0)^2 + Y(0)^2 + N \cdot C_1} \right\} \cdot \\ &\times \left\{ 1 - \frac{\sum_{k=1}^{N-1} (X(k) - Y(k))^2}{\sum_{k=1}^{N-1} \left(X(k)^2 + Y(k)^2 \right) + N \cdot C_2} \right\},\end{aligned} \tag{23}$$

where C_1 and C_2 are the same constants as in Eq. (1); N is the number of DCT coefficients; $X(k)$ and $Y(k)$ represent the DCT coefficients of x and y, respectively. From the equation, the DCT domain SSIM can be characterized by the product of DC ($k = 0$) and AC ($k = 1, 2, \ldots, N-1$) similarities. Inspired by this, the RR SSIM [38] is also calculated as the product of a DC and an AC components. More specifically, each frame is first divided into 4×4 non-overlapping blocks, followed by a block DCT transform. Next, all DCT coefficients of the same frequency are grouped into one subband, which results in $N = 16$ subbands. Finally, the RR distortion measure is defined as

$$M_{RR} = \left(1 - \frac{\text{MSE}_0}{2\sigma_0^2 + C_1}\right) \cdot \left(1 - \frac{1}{N-1} \sum_{k=1}^{N-1} \frac{\text{MSE}_k}{2\sigma_k^2 + C_2}\right), \tag{24}$$

where σ_k is the standard derivation of the kth subband and MSE_k denote the MSE between original and distorted frames

TABLE I
ERRORS IN PARAMETER ESTIMATION OF SSIM-Q AND R-Q MODELS

Configurations	Prediction errors of Λ			Prediction errors of Λ_k			Prediction errors of Ω		
	PREV1	PREV2	PREV3	PREV1	PREV2	PREV3	PREV1	PREV2	PREV3
GOP=1	0.00763	0.00843	0.01011	0.00873	0.00844	0.00901	0.01951	0.02081	0.02249
GOP=2	0.01031	0.01113	0.01271	0.00980	0.00958	0.01051	0.03044	0.02968	0.03015
GOP=4	0.01193	0.01217	0.01393	0.01060	0.01046	0.01149	0.03467	0.03245	0.03291
GOP=8	0.01220	0.01265	0.01422	0.01112	0.01112	0.01215	0.03716	0.03557	0.03562
Average	0.01052	0.01110	0.01274	0.01006	0.00990	0.01079	0.03044	0.02962	0.03029

in the kth subband. Due to the fact that the DCT coefficients can be modeled by Laplacian distributions [46]

$$f_{Lap}(x) = \frac{\Lambda}{2} \cdot e^{-\Lambda \cdot |x|}, \tag{25}$$

MSE_k can be estimated by

$$\begin{aligned}\mathrm{MSE}_k &= \int_{-(Q-\gamma Q)}^{Q-\gamma Q} x_k^2 f_{Lap}(x_k)dx_k \\ &\quad + 2\sum_{n=1}^{\infty} \int_{nQ-\gamma Q}^{(n+1)Q-\gamma Q} (x_k - nQ)^2 f_{Lap}(x_k)dx_k \\ &= \frac{2}{\Lambda_k^2} + \left[(1-2\gamma)Q^2 + \frac{2Q}{\Lambda_k}\right]\frac{e^{\gamma \Lambda_k Q}}{1-e^{\Lambda_k Q}},\end{aligned} \tag{26}$$

where γ is the rounding offset in quantization.

Experiments show that there exists a nearly perfect linear relationship between M_{RR} and SSIM [38]. Specifically, by Eq. (24), $M_{RR} = 1$ when there is no distortion; and in such a case, the value of SSIM is also 1. Hence, SSIM can be predicted by

$$\mathrm{SSIM} = (1-\phi) + \phi \cdot M_{RR}, \tag{27}$$

where ϕ is a prediction parameter.

Our R-Q model is based on the entropy model in [16],

$$R = H \cdot e^{\mu \Lambda Q + \nu}, \tag{28}$$

where μ and ν are constants, and Λ is the Laplacian parameter of the coding residuals. Let P_S denote the probability of skipped blocks, $\Phi = 1 - e^{-(1-\gamma)\Lambda Q}$, $\Psi = 1 - e^{-\Lambda Q}$, and $\Omega = \frac{P_S}{\Phi}$, the entropy H can be derived as

$$\begin{aligned}H = \frac{1}{\ln 2}\Bigg\{&(1-\Omega\Phi)\cdot \ln(1-\Omega\Phi) \\ &- (1-\Omega)\Phi \cdot \ln[(1-\Omega)\Phi] \\ &+ (1-\Phi)\cdot\left[\ln\frac{2}{\Psi} + \Lambda Q\left(\frac{1}{\Psi} - \gamma\right)\right]\Bigg\}.\end{aligned} \tag{29}$$

Specifically, $H = 0$ when $P_S = 1$.

C. Implementation Issues

As shown in Fig. 5, SVC supports an HBP structure, in which the Qps, frame distances and prediction frames are different among all temporal layers T0, T1, T2 and T3. Therefore, we predict the model parameters, including the coding residual parameter Λ, the distortion parameter Λ_k and the entropy parameter Ω, from the nearest pre-coded frames of the same temporal layer. To decide how many pre-coded frames are required, we compare the average prediction error of PREV1 (one pre-coded frame used), PREV2 (the average of two pre-coded frames) and PREV3 (the average of three pre-coded frames). 9 sequences (including 3 CIF, 3 4CIF and 3 HD sequences), 4 GOP sizes (1, 2, 4, 8), 4 Qp settings (10, 20, 30, 40) and 33 frames are tested with Context-Based Adaptive Binary Arithmetic Coding (CABAC). The comparison results are summarized in Table I, which suggests that for an HBP structure, PREV1, PREV2 and PREV2 can achieve the best prediction accuracy for Λ, Λ_k and Ω, respectively. As a result, for a frame with temporal layer T and CGS layer L, these parameters are predicted as

$$\begin{aligned}\hat{\Lambda}(L,T;t) &= \Lambda(L,T;t-1), \\ \hat{\Lambda}_k(L,T;t) &= \frac{1}{2}[\Lambda_k(L,T;t-1) + \Lambda_k(L,T;t-2)], \\ \hat{\Omega}(L,T;t) &= \frac{1}{2}[\Omega(L,T;t-1) + \Omega(L,T;t-2)],\end{aligned} \tag{30}$$

where t represents the frame index.

In the SSIM-Q model, the parameter ϕ can be obtained after a frame is coded based on SSIM, M_{RR}, and Eq. (27). To predict ϕ before coding a frame, the average value of those in the nearest two pre-coded frames is used:

$$\hat{\phi}(L,T;t) = \frac{1}{2}[\phi(L,T;t-1) + \phi(L,T;t-2)]. \tag{31}$$

In the R-Q model, the header bits are also taken into consideration for a complete model, which results in different μ and ν from [16]. Besides, large GOP sizes and different temporal layers also change the R-H relationship in Eq. (28). In this work, we follow the parameters in [38] for I and P frames and keep the corresponding LMs fixed; for B frames, we train the two parameters with a set of 4CIF sequences and obtain

$$\begin{aligned}\mu &= 0.05 \cdot T - 0.25, \\ \nu &= 0.08 \cdot G - 0.5 \cdot T - 1.35,\end{aligned} \tag{32}$$

where G and T denote the GOP size and temporal layer, respectively.

Finally, considering different CGS and temporal layers, Eq. (19) is rewritten as

$$\lambda_{L,T} = \Theta_L \cdot \frac{\partial \mathrm{SSIM}_{L,T}}{\partial R_{L,T}} = \Theta_L \cdot \frac{\frac{\partial \mathrm{SSIM}_{L,T}}{\partial Q_{L,T}}}{\frac{\partial R_{L,T}}{\partial Q_{L,T}}}. \tag{33}$$

For each CGS and temporal layer, the partial differentials are derived from the SSIM-Q model in Eqs. (24), (26), and (27),

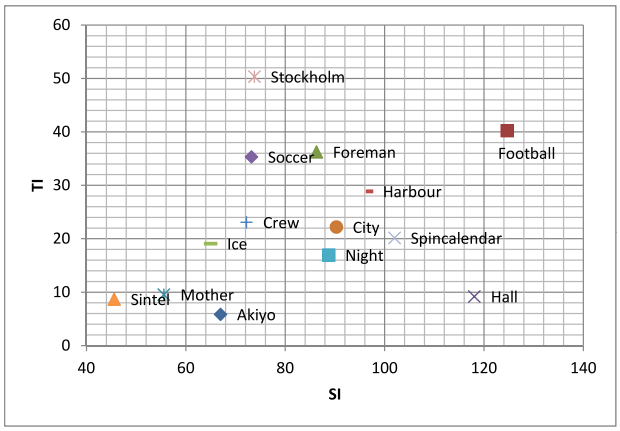


Fig. 6. Illustration of SI and TI values of all tested sequences.

and the R-Q model in Eqs. (28) and (29), with the parameters predicted by Eqs. (30), (31), and (32). In the first several frames, the LM is initialized by λ_{HR} [38], as

$$\lambda_{HR} = \xi \cdot Q^2 - \zeta \cdot Q^4, \tag{34}$$

where

$$\xi = \begin{cases} 2.1 \times 10^{-4} & \text{B frame,} \\ 7 \times 10^{-5} & \text{otherwise,} \end{cases}$$
$$\zeta = \begin{cases} 1.5 \times 10^{-9} & \text{B frame,} \\ 5 \times 10^{-10} & \text{otherwise.} \end{cases} \tag{35}$$

After a frame is coded, the average SSIM index is calculated as a weighted sum of SSIM indices of its Y, Cb and Cr components:

$$\text{SSIM} = W_Y \cdot \text{SSIM}_Y + W_{Cb} \cdot \text{SSIM}_{Cb} + W_{Cr} \cdot \text{SSIM}_{Cr}, \tag{36}$$

where $W_Y = 0.8$, $W_{Cb} = W_{Cr} = 0.1$, as in [47].

IV. Experiments

To validate the proposed CGS coding scheme, we implement the proposed framework on the SVC reference software JSVM 9.19.14. Five CIF sequences (*Akiyo*, *Football*, *Foreman*, *Hall* and *Mother*), five 4CIF sequences (*City*, *Crew*, *Harbor*, *Ice* and *Soccer*) and four HD sequences (*Night*, *Sintel*, *Spincalendar* and *Stockholm*) are tested with YCbCr 4:2:0 format, which cover a large range of Spatial Information (SI) and Temporal Information (TI) [48] values, as shown in Fig. 6. To examine the robustness of our algorithm, we test both two-CGS layer and three-CGS layer settings, with four GOP structures (GOP = 1, 2, 4, and 8).

The simulation environment is summarized as follows: 1) High and scalable high profiles are used in BL and EL(s), respectively; 2) Intra period is -1 (i.e., only the very first frame is totally intra coded), and 241 frames are coded for all GOP sizes; 3) The number of reference frames is 2; 4) Fast search is enabled with Hadamard function based sub-pixel search; 5) The search ranges are ±32, ±64 and ±64 for CIF, 4CIF and HD sequences, respectively; 6) CABAC coding mode is enabled due to its efficiency; 7) The other parameters are set as the defaults of the reference software.

To observe the improvement of the proposed scheme, we also implemented the single layer RDO method [38] and the conventional MSE-based RDO scheme for comparison. For each sequence with a specific Qp, the overall R-S performance is measured by weighted sums of SSIMs and BRs for all CGS layers, as $\overline{S} = \sum_{L=0}^{L_{max}} w_L \text{SSIM}_L$ and $\overline{R} = \sum_{L=0}^{L_{max}} w_L R_L$, respectively. To compare the average R-S performance with multiple Qps, we use the Bjontegaard average BR increase (BDBR, %) [49], but with weighted sums, $\overline{S}$ and $\overline{R}$, instead of conventionally adopted PSNR and BR.

A. Simulation With Two CGS Layers

The coding performance of the proposed scheme is first examined with two CGS layers (BL & EL). Four groups of Qp settings are employed, which are (20, 15), (25, 20), (30, 25) and (35, 30) where the first Qp in each group is for BL (layer ID $L = 0$) and the second is for EL ($L = 1$), respectively. In such a case, the maximum Qp for all frames is 38, when temporal layer is 3, GOP size is 8, and BL Qp equals 35; the minimum Qp for all frames is 13, when temporal layer is 0, GOP size is 8, and EL Qp equals 15. Hence, our test has covered a large range of Qp values from low BR end to high BR end.

A significant feature of the proposed RDO scheme is that it allows for arbitrary user-defined weights assigned to different layers. To examine this, three groups of weights are used, which are $w(1{:}2)$ (i.e., $w_0 = 1/3$, $w_1 = 2/3$) for increasing weights, $w(2{:}1)$ (i.e., $w_0 = 2/3$, $w_1 = 1/3$) for decreasing weights, and $w(1{:}1)$ (i.e., $w_0 = 1/2$, $w_1 = 1/2$) for uniform weight, respectively.

The proposed multiple layer RDO scheme is compared with single layer RDO method [38]. The BDBRs for all sequences, GOP sizes and user-defined weights are shown in Table II. The maximum BDBR reduction is 6.14% when coding *Soccer* with $w_0 : w_1 = 1 : 2$ and GOP = 4. On average, the proposed scheme achieves up to 3.29% BDBR reduction while keeping the same SSIM quality as in [38], with $w_0 : w_1 = 1 : 2$ and GOP = 2. Compared with single layer RDO method, the proposed scheme achieves better overall R-S performance when $w_0{:}w_1 = 1 : 1$ and $w_0{:}w_1 = 1 : 2$, which corresponds to the case that EL has similar or larger weight than BL.

To justify the improvement of R-S performance compared with the conventional MSE-based RDO method, the BDBRs between the proposed scheme and the original CGS encoder are given in Table III. The maximum BDBR reduction is 15.60% when coding *Sintel* with $w_0 : w_1 = 1 : 2$, GOP = 1; and the maximum average BDBR reduction is 5.83% with $w_0 : w_1 = 2 : 1$ and GOP = 1. In general, smaller GOP size is more likely to result in better R-S performance. The average BDBR reduction of the GOP sizes 1, 2, 4, and 8 are 5.36%, 3.43%, 3.46% and 2.87%, respectively. Some sequences may have bit rate increase with large GOP sizes, such as *Foreman*, *Harbor* and *Spincalendar*, because SSIM-Q and R-Q parameters prediction cannot always achieve good performances for large GOP sizes where the intervals between frames are

TABLE II
COMPARISON OF THE PROPOSED SCHEME WITH SINGLE LAYER RDO METHOD [38] FOR TWO-LAYER CGS

Sequences	GOP=1			GOP=2			GOP=4			GOP=8		
	w(1:1)	w(1:2)	w(2:1)	w(1:1)	w(1:2)	w(2:1)	w(1:1)	w(1:2)	w(2:1)	w(1:1)	w(1:2)	w(2:1)
Akiyo	-2.82	-4.44	-1.31	-3.14	-6.04	-1.43	-2.49	-5.09	-0.64	-2.05	-4.08	-1.28
Football	-0.85	-1.00	-0.39	-0.47	-0.53	-0.23	-0.38	-0.37	-0.17	-0.71	-0.93	-0.30
Foreman	-2.08	-3.08	-1.19	-2.21	-2.95	-1.31	-1.74	-2.49	-0.93	-1.14	-1.88	-0.38
Hall	-1.80	-1.80	-1.13	-1.52	-1.59	-0.70	-1.31	-2.05	-1.24	-1.62	-1.71	-1.28
Mother	-1.47	-2.42	-0.67	-1.06	-1.65	-0.71	-0.72	-0.81	-0.71	-0.25	-0.38	0.03
City	-3.31	-4.34	-2.09	-4.22	-5.78	-2.56	-3.52	-4.98	-2.21	-3.31	-4.19	-2.16
Crew	-0.75	-0.96	-0.43	-0.72	-1.10	-0.31	-0.65	-0.99	-0.37	-0.74	-0.87	-0.35
Harbour	-3.47	-4.60	-2.09	-3.32	-4.80	-1.96	-2.51	-3.86	-1.46	-1.98	-3.04	-1.13
Ice	-2.02	-3.40	-1.12	-2.03	-2.98	-0.64	-1.33	-2.66	-0.54	-1.04	-1.84	-0.64
Soccer	-1.53	-1.94	-0.87	-1.82	-2.12	-1.08	-4.93	-6.14	-3.34	-1.06	-1.03	-0.57
Night	-1.73	-2.31	-0.97	-2.05	-3.10	-1.04	-1.83	-2.74	-0.85	-1.44	-2.31	-0.80
Sintel	-2.92	-5.50	-1.32	-2.33	-4.68	-1.02	-1.99	-3.62	-0.87	-1.24	-2.85	-0.84
Spincalendar	-3.12	-4.54	-1.85	-3.57	-5.57	-2.03	-3.24	-5.50	-1.70	-3.02	-5.01	-1.54
Stockholm	-1.93	-2.11	-1.31	-2.36	-3.12	-1.39	-2.08	-2.71	-1.30	-1.66	-2.05	-0.96
Average	-2.13	-3.03	-1.20	-2.20	-3.29	-1.17	-2.05	-3.14	-1.17	-1.52	-2.30	-0.87

TABLE III
COMPARISON OF THE PROPOSED SCHEME WITH CONVENTIONAL RDO METHOD FOR TWO-LAYER CGS

Sequences	GOP=1			GOP=2			GOP=4			GOP=8		
	w(1:1)	w(1:2)	w(2:1)	w(1:1)	w(1:2)	w(2:1)	w(1:1)	w(1:2)	w(2:1)	w(1:1)	w(1:2)	w(2:1)
Akiyo	-5.56	-5.21	-7.05	0.74	-0.82	0.28	-0.39	-1.56	-0.56	-0.63	-1.42	-1.49
Football	-3.82	-3.21	-4.13	-3.87	-3.27	-4.23	-3.20	-2.79	-3.43	-1.89	-1.77	-1.99
Foreman	-1.73	-1.55	-2.31	-0.37	-0.04	-0.68	-0.44	-0.24	-0.80	0.53	0.65	0.16
Hall	-6.41	-5.50	-8.26	-4.99	-4.17	-6.06	-4.66	-5.00	-5.93	-4.38	-4.36	-5.37
Mother	-7.49	-6.93	-7.68	-4.87	-4.98	-4.52	-3.08	-3.43	-2.57	-2.23	-3.22	-1.36
City	-5.16	-4.92	-6.17	-1.88	-2.03	-2.19	-1.11	-1.08	-1.84	-1.67	-0.75	-2.91
Crew	-3.30	-2.42	-3.98	-3.36	-2.83	-3.75	-3.90	-3.62	-4.16	-4.28	-3.79	-4.45
Harbour	1.84	0.20	2.65	2.49	1.12	3.01	1.11	0.36	1.07	1.38	0.82	1.40
Ice	-11.25	-10.22	-12.56	-10.65	-9.82	-10.82	-9.58	-9.50	-9.90	-8.16	-7.80	-8.71
Soccer	-8.03	-6.58	-9.08	-7.42	-5.88	-8.28	-9.31	-8.59	-9.77	-6.48	-4.66	-7.74
Night	-5.34	-4.72	-6.34	-2.32	-2.29	-2.79	-2.31	-2.23	-2.64	-2.48	-2.32	-3.14
Sintel	-14.16	-15.60	-12.96	-11.52	-14.06	-9.76	-10.44	-12.63	-8.74	-8.91	-11.12	-7.72
Spincalendar	-0.51	-0.94	-0.81	2.12	1.17	2.07	2.07	0.87	1.96	2.20	1.41	1.68
Stockholm	-2.88	-2.03	-3.01	-1.01	-0.94	-0.51	-1.25	-1.21	-0.84	-1.44	-1.19	-0.84
Average	-5.27	-4.97	-5.83	-3.35	-3.49	-3.45	-3.32	-3.62	-3.44	-2.75	-2.82	-3.03

also large. How to accurately predict SSIM-Q and R-Q parameters between frames with large distance will need to be studied. In particular, the above three sequences have large SI (i.e., complex texture) and/or large TI (i.e., fast motion), as shown in Fig. 6, which increases the probabilities of inaccurate parameter predictions.

For some sequences, the BDBRs in Table III are larger than those corresponding values in Table II, e.g., when encoding *Foreman* with $w_0 : w_1 = 1 : 1$ and GOP $= 1$. In other words, the single layer SSIM-based RDO method [38] may not always achieve better overall R-S performance than the conventional RDO scheme. This may be due to several reasons. First, the RDO method reported in [38] works better for P frames than B frames, leading to limited coding performance in an HBP structure. Second, because of the large frame distances in hierarchical layers, there may exist large errors in the prediction of parameters, as discussed earlier. Third, this method is designed for single layer encoding and may not work well for EL, when the frames can be predicted from both EL and BL.

B. Simulation With Three CGS Layers

To further validate the coding performance of the proposed scheme, we test it with three CGS layers (BL & EL1 & EL2). We use four groups of Qp settings, given by (20, 15, 10), (25, 20, 15), (30, 25, 20) and (35, 30, 25), respectively, where the first, second and third Qps are for BL ($L = 0$), EL1 ($L = 1$) and EL2 ($L = 2$), respectively. The range of Qps for all frames is from 8 to 38. Three groups of layer weights are tested: w(1:2:3) (i.e., $w_0 = 1/6$, $w_1 = 1/3$, $w_2 = 1/2$) for increasing weights, w(3:2:1) (i.e., $w_0 = 1/2$, $w_1 = 1/3$, $w_2 = 1/6$) for decreasing weights, and w(1:1:1) (i.e., $w_0 = 1/3$, $w_1 = 1/3$, $w_2 = 1/3$) for uniform weight, respectively.

The coding performances in terms of BDBR, are given in Tables IV and V, respectively, with comparisons to single layer SSIM-based RDO [38] and conventional RDO methods. The proposed scheme achieves a good performance on average and also for most individual sequences, showing the effectiveness and robustness of the proposed scheme. Compared with [38] and the original CGS encoder, the proposed scheme can achieve up to 6.45% and 14.23% BDBR

TABLE IV
COMPARISON OF THE PROPOSED SCHEME WITH SINGLE LAYER RDO METHOD [38] FOR THREE-LAYER CGS

Sequences	GOP=1			GOP=2			GOP=4			GOP=8		
	w(1:1:1)	w(1:2:3)	w(3:2:1)	w(1:1:1)	w(1:2:3)	w(3:2:1)	w(1:1:1)	w(1:2:3)	w(3:2:1)	w(1:1:1)	w(1:2:3)	w(3:2:1)
Akiyo	-2.93	-4.85	-1.82	-4.08	-6.45	-2.49	-2.88	-5.22	-1.99	-2.84	-4.87	-1.60
Football	-1.17	-1.63	-0.84	-0.86	-0.81	-0.48	-0.85	-1.00	-0.23	-0.99	-1.69	-0.43
Foreman	-2.37	-2.55	-1.72	-2.19	-2.60	-1.91	-1.85	-2.19	-1.16	-1.39	-2.03	-0.92
Hall	-3.25	-2.72	-2.25	-2.89	-2.28	-1.93	-2.80	-2.43	-2.01	-2.75	-2.09	-1.90
Mother	-2.08	-2.48	-1.55	-1.98	-2.21	-1.39	-1.51	-1.37	-1.04	-0.69	-0.69	0.00
City	-4.35	-4.76	-3.15	-4.78	-5.67	-3.58	-4.45	-5.42	-3.32	-4.14	-5.38	-3.15
Crew	-1.03	-1.03	-0.83	-1.01	-1.12	-0.72	-1.14	-1.42	-0.83	-1.33	-1.69	-0.90
Harbour	-4.63	-5.37	-3.31	-4.18	-5.11	-3.01	-3.26	-4.23	-2.32	-2.56	-3.51	-1.81
Ice	-2.98	-3.67	-2.04	-2.49	-3.37	-1.59	-2.29	-3.15	-1.43	-1.85	-2.95	-1.13
Soccer	-2.53	-2.67	-1.73	-2.51	-2.71	-1.80	-2.28	-2.45	-1.65	-1.90	-2.13	-1.27
Night	-2.42	-2.86	-1.70	-2.56	-3.32	-1.80	-2.33	-3.04	-1.63	-2.04	-2.75	-1.36
Sintel	-3.97	-6.80	-2.41	-3.63	-5.44	-1.96	-2.80	-4.45	-1.61	-2.37	-3.38	-1.78
Spincalendar	-3.83	-4.94	-2.76	-4.03	-5.57	-2.93	-4.01	-5.62	-2.84	-4.05	-5.83	-2.67
Stockholm	-2.29	-2.31	-1.80	-2.50	-2.82	-2.00	-2.37	-2.81	-1.90	-2.35	-2.76	-1.78
Average	-2.85	-3.47	-1.99	-2.83	-3.53	-1.97	-2.49	-3.20	-1.71	-2.23	-2.98	-1.48

TABLE V
COMPARISON OF THE PROPOSED SCHEME WITH CONVENTIONAL RDO METHOD FOR THREE-LAYER CGS

Sequences	GOP=1			GOP=2			GOP=4			GOP=8		
	w(1:1:1)	w(1:2:3)	w(3:2:1)	w(1:1:1)	w(1:2:3)	w(3:2:1)	w(1:1:1)	w(1:2:3)	w(3:2:1)	w(1:1:1)	w(1:2:3)	w(3:2:1)
Akiyo	-8.37	-12.92	-6.09	-4.65	-10.30	-1.05	-4.06	-9.43	-1.60	-4.07	-9.11	-1.28
Football	-4.71	-5.20	-4.54	-4.60	-4.13	-4.61	-3.36	-3.12	-3.33	-1.49	-1.80	-1.54
Foreman	-2.08	-2.94	-1.77	-0.98	-1.97	-0.84	-0.69	-1.32	-0.28	0.34	-0.67	0.52
Hall	-9.04	-9.77	-8.04	-7.98	-8.56	-6.86	-7.96	-9.11	-6.67	-7.31	-8.43	-5.67
Mother	-2.38	-2.49	-3.57	-2.11	-2.01	-2.73	-2.57	-2.17	-2.62	-3.13	-3.15	-2.38
City	-4.43	-5.50	-3.85	-3.12	-4.45	-1.98	-2.69	-4.15	-1.56	-2.65	-4.13	-1.94
Crew	-2.58	-2.08	-3.12	-2.55	-2.28	-2.88	-2.29	-2.53	-2.54	-1.97	-2.35	-2.24
Harbour	1.36	-1.42	3.29	1.05	-1.20	2.85	0.15	-1.28	1.16	0.62	-0.81	1.48
Ice	-7.23	-8.23	-7.57	-6.41	-7.30	-6.79	-5.27	-6.20	-5.69	-3.58	-4.91	-4.09
Soccer	-7.01	-6.57	-7.31	-6.75	-6.01	-7.11	-6.21	-5.47	-6.78	-5.04	-4.44	-5.75
Night	-4.01	-4.83	-4.00	-2.16	-3.48	-1.62	-1.81	-3.10	-1.33	-1.74	-2.71	-1.49
Sintel	-12.00	-14.23	-11.73	-9.80	-11.59	-8.89	-8.26	-10.20	-7.55	-6.09	-7.11	-6.42
Spincalendar	-2.03	-3.80	-0.91	-0.01	-2.42	1.51	0.98	-1.76	2.58	1.64	-1.43	3.46
Stockholm	-3.27	-3.10	-3.08	-1.91	-2.36	-1.39	-1.01	-1.96	-0.44	-0.09	-1.29	0.68
Average	-4.84	-5.93	-4.45	-3.71	-4.86	-3.03	-3.22	-4.42	-2.62	-2.47	-3.74	-1.91

TABLE VI
VISUAL TEST OF VIDEOS WITH SIMILAR SSIM. SSIM, PSNR, BR, AND MOS REPRESENTS $\overline{S}$, $\overline{P}$, $\overline{R}$, AND $\overline{MOS}$, RESPECTIVELY

Sequences		$w(1:1:1)$				$w(1:2:3)$				$w(3:2:1)$			
		SSIM	PSNR	BR	MOS	SSIM	PSNR	BR	MOS	SSIM	PSNR	BR	MOS
Mother, CIF	Conventional	0.9573	40.90	285.25	6.9444	0.9642	41.91	364.98	7.2222	0.9496	39.83	198.56	6.1806
	Proposed	0.9572	40.57	269.88	6.3055	0.9646	41.68	345.82	7.4028	0.9497	39.47	192.83	5.8611
Soccer, 4CIF	Conventional	0.9103	36.47	1218.52	7.1667	0.9185	36.96	1411.57	7.0741	0.9020	35.99	1025.47	7.2593
	Proposed	0.9096	36.03	1109.09	6.7778	0.9180	36.59	1280.65	7.2222	0.9013	35.47	931.92	7.3704
Sintel, HD	Conventional	0.9916	46.82	701.79	4.6944	0.9925	47.78	803.03	4.5139	0.9907	45.94	510.08	3.3611
	Proposed	0.9916	46.43	615.41	3.8056	0.9925	47.38	774.08	5.2917	0.9907	45.48	458.81	3.1806

reductions with the same SSIM quality, respectively. On average, the proposed scheme can achieve up to 3.53% and 5.93% BDBR reductions, respectively, whilst keeping the same SSIM quality as compared to the single layer and conventional RDO method.

C. Subjective Test

To further verify the performance of the proposed scheme, we carry out subjective test in addition to BDBR shown in Sections IV-A and IV-B. Due to the enormous variations in video sequences, Qps, layer weights and other settings, subjective test of all cases is impossible. Therefore, subjective verification focuses on the representative cases. The proposed subjective test consists of two parts. In the first part, we examine the visual qualities of video sequences with similar weighted sum of SSIMs, $\overline{S}$. Table VI shows some two-layer and three-layer cases where the conventional and proposed methods achieve similar $\overline{S}$. The weighted sum of PSNR values, $\overline{P} = \sum_{L=0}^{L_{max}} w_L \mathrm{PSNR}_L$, is also listed for reference. Sequences with three resolutions and three types of weights are tested. On average, the weighted sum of BRs, $\overline{R}$, of the proposed scheme is reduced by 7.43% as compared with the conventional algorithm.

We perform subjective test to examine the visual quality score of each sequence. In the test, there are a total of

TABLE VII
VISUAL PREFERENCE OF VIDEOS WITH SIMILAR BR. SSIM, BR, PSNR, AND VP REPRESENTS $\overline{S}$, $\overline{R}$, $\overline{P}$, AND $\overline{VP}$, RESPECTIVELY

Sequences		SSIM	PSNR	BR	VP
Crew, 4CIF, $w(1:1)$	Conventional	0.9072	36.92	1317.68	42.31
	Proposed	0.9102	36.72	1210.68	57.69
Soccer, 4CIF, $w(1:1:1)$	Conventional	0.9210	37.59	2288.90	37.82
	Proposed	0.9277	37.59	2280.93	62.18

41 sequences, which includes all layers of sequences in Table VI except some duplicated sequences (e.g., the three conventional sequences of *Soccer* in Table VI are the same, just with different layer weights). We put the 41 stimuli in a random order and presented them to 15 subjects, who were asked to mark each sequence with a subjective score from 0 to 10, where 0 and 10 represent video sequences with totally distorted and perfect quality, respectively. To increase the reliability of human scores, we repeated 3 stimuli twice, which results in 44 stimuli totally, and use the maximum absolute score difference of identical stimuli (MaxASD) to eliminate unreliable scores. According to observation of the final scores, we excluded invalid results where MaxASD≥4.

By averaging all the remaining scores, we obtain the Mean Opinion Score (MOS) and show all the results in Table VI, where $\overline{MOS} = \sum_{L=0}^{L_{max}} w_L \mathrm{MOS}_L$. The ranges of $\overline{S}$ and $\overline{MOS}$ are 0.9013∼0.9925 and 3.1806∼7.4028, respectively. On the other hand, the average and maximum absolute differences between $\overline{MOS}$s of conventional and our methods are 0.4038 and 0.8889 (*Sintel*, w(1:1:1)), respectively, which are relatively small considering the large range of MOS for a small range of SSIM. In addition, as a reference, among all valid results, the mean absolute score difference of the 3 identical stimuli is 1.4444. Hence, we conclude that, in each group of the two tables, the proposed scheme reduces BR of coded videos without any visual quality changes.

Another useful observation is that, with similar SSIM and MOS values, the PSNR values often differ significantly among them. For all sequences in Table VI, the largest PSNR difference is 0.52 dB (*Soccer, 4CIF* w(2:1)), while the SSIM and MOS differences are 0.0007 and 0.1111, respectively. This fact can be considered as another evidence that SSIM is more consistent with human visual perception as compared with PSNR. In general, the video sequences created by the proposed scheme have lower PSNR, but their SSIM and MOS values remain about the same as those generated by the conventional method.

In the second part of the test, we examine the video sequences with similar weighted sum of coding BRs, $\overline{R}$. To highlight the differences, we repeat a few number of sequences several times and compare the sequences via a two-alternative-forced choice (2AFC) method. We work with five groups of sequences (two layers of *Crew* and three layers of *Soccer*) and in each group, there are two sequences coded by conventional and the proposed methods, respectively. We repeated the five groups by four times and produced 20 randomly sorted stimuli for 13 subjects. In each stimuli, a subject was asked to choose one of the two sequences he/she thought to have better quality, where the two sequences were also randomly ordered. The results are summarized in Table VII, where PSNR is also included for reference; VP indicates a Video Preference Score which indicates the percentage when the subjects think this video is of better quality; $\overline{VP} = \sum_{L=0}^{L_{max}} w_L \mathrm{VP}_L$. On average, the $\overline{VP}$s of the proposed scheme are 57.69% and 62.18% for *Crew* and *Soccer*, respectively, which indicates that more often the subjects think the sequences coded by the proposed scheme are of better quality than those coded by the conventional scheme. The human scores are consistent with SSIM which also demonstrates that the proposed scheme achieves better quality, even though with similar or lower PSNR than the conventional method. In conclusion, the proposed scheme can improve the visual quality of the original algorithm, either from SSIM or VP, while maintaining almost the same coding bits.

D. Complexity Analysis

The inter-layer R-S performance improvement mainly comes from Eq. (22), where its computational overhead is negligible considering the high complexity of RDO and ME in video coding. A more relevant issue that may increase computational complexity is the calculation of SSIM in RDO process. It was shown in [50] that the SSIM computation complexity is about 5% of that of the whole mode decision process. In our method, the SSIM calculation is based on the scheme proposed by Wang *et al.* [38], where the SSIM computation was demonstrated to increase the computational complexity by 6.3% on average.

To further reduce the complexity, we may utilize the fast SSIM algorithms [51] and [52] or simplify the SSIM calculation by modifying the calculation window. In [53], the SSIM is calculated based on 4×4 non-overlapping blocks. In [26], a block-based SSIM calculation approach was developed, which could reduce the computational time by skipping SSIM calculation on some pixels. In hardware platform, the SSIM calculation can be further boosted with parallel calculations of average and standard variance values.

V. CONCLUSION

We propose an SSIM-based CGS coding scheme, which facilitates different user-defined weights for different CGS layers. Based on investigations of inter-layer R-S relationships, an inter-layer R-S dependency model is proposed. By incorporating the dependency model and SSIM-Q, R-Q models, an RDO scheme is devised to maximize the weighted sum of R-S performances of all CGS layers. The experimental results demonstrate the effectiveness and robustness of the proposed scheme, which is superior to both single layer SSIM-based RDO method and conventional MSE-based RDO scheme.

REFERENCES

[1] *Advanced Video Coding for Generic Audiovisual Services*, ISO/IEC Standard 14496-10(E), ITU-T Recommendation H.264(E), Mar. 2005.

[2] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm, and T. Wiegand, *High Efficiency Video Coding (HEVC) Text Specification Draft 8*, JCTVC Document JCTVC-J1003, Jul. 2012.

[3] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien, *Joint Draft 11: Scalable Video Coding*, JVT Document JVT-X201, Jul. 2007.

[4] K. Mueller, P. Merkel, A. Smolic, and T. Wiegand, *Multiview Coding Using AVC*, ISO/IEC JTC1/SC29/WG11 Document M12945, 2006.

[5] M. Wien, "Variable block-size transforms for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 604–613, Jul. 2003.

[6] T. Wiegand and B. Girod, *Multi-Frame Motion-Compensated Prediction for Video Transmission*. London, U.K.: Kluwer Academic, 2001.

[7] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 620–636, Jul. 2003.

[8] G. Bjontegaard and K. Lillevold, *Context-Adaptive VLC Coding of Coefficients*, JVT Document JVT-C028, May 2002.

[9] P. List, A. Joch, J. Lainema, G. Bjontegaard, and M. Karczewicz, "Adaptive deblocking filter," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 614–619, Jul. 2003.

[10] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and compression of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688–703, Jul. 2003.

[11] W. J. Han *et al.*, "Improved video compression efficiency through flexible unit representation and corresponding extension of coding tools," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1709–1720, Dec. 2010.

[12] 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.

[13] G. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.

[14] Y. Yang and S. S. Hemami, "Generalized rate-distortion optimization for motion-compensated video coders," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 942–955, Sep. 2000.

[15] E. H. Yang and X. Yu, "Rate distortion optimization for H.264 inter-frame coding: A general frame work and algorithms," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1774–1784, Jul. 2007.

[16] 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. 19, no. 2, pp. 193–205, Feb. 2009.

[17] X. Zhao, L. Zhang, S. Ma, and W. Gao, "Video coding with rate-distortion optimized transform," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 1, pp. 138–151, Jan. 2012.

[18] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[19] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[20] 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.

[21] K. Zeng, A. Rehman, J. Wang, and Z. Wang, "From H.264 to HEVC: Coding gain predicted by objective video quality assessment models," in *Proc. 7th Int. Workshop Video Process. Qual. Metr. Consumer Electron. (VPQM)*, Scottsdale, AZ, USA, Jan./Feb. 2013, pp. 42–46.

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

[23] 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.

[24] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3449–3451, Nov. 2006.

[25] D. Brunet, E. R. Vrscay, and Z. Wang, "On the mathematical properties of the structural similarity index," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1488–1499, Apr. 2012.

[26] T. Zhao, K. Zeng, A. Rehman, and Z. Wang, "On the use of SSIM in HEVC," in *Proc. 47th IEEE Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, 2013, pp. 1107–1111.

[27] Z.-Y. Mai, C.-L. Yang, L.-M. Po, and S.-L. Xie, "A new rate-distortion optimization using structural information in H.264 I-frame encoder," in *Proc. 7th Int. Conf. Adv. Concepts Intell. Vis. Syst. (ACIVS)*, Antwerp, Belgium, 2005, pp. 435–441.

[28] 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 *Proc. IEEE Int. Conf. Syst. Man Cybern. (SMC)*, vol. 3. Waikoloa, HI, USA, Oct. 2005, pp. 2673–2678.

[29] B. H. K. Aswathappa and K. R. Rao, "Rate-distortion optimization using structural information in H.264 strictly intra-frame encoder," in *Proc. 42nd Southeast. Symp. Syst. Theory (SSST)*, Tyler, TX, USA, Mar. 2010, pp. 367–370.

[30] 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 *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2006, pp. 913–916.

[31] C.-L. Yang, H.-X. Wang, and L.-M. Po, "Improved inter prediction based on structural similarity in H.264," in *Proc. Int. Conf. Signal Process. Commun. (ICSPC)*, Dubai, UAE, Nov. 2007, pp. 340–343.

[32] C.-L. Yang, R.-K. Leung, L.-M. Po, and Z.-Y. Mai, "An SSIM-optimal H.264/AVC inter frame encoder," in *Proc. Int. Conf. Intell. Comput. Intell. Syst. (ICIS)*, vol. 4. Shanghai, China, Nov. 2009, pp. 291–295.

[33] T.-S. Ou, Y.-H. Huang, and H. H. Chen, "A perceptual-based approach to bit allocation for H.264 encoder," *SPIE Vis. Commun. Image Process.*, vol. 7744, Jul. 2010, Art. ID 77441B.

[34] Y.-H. Huang, T.-S. Ou, and H. H. Chen, "Perceptual-based coding mode decision," in *Proc. Int. Symp. Circuits Syst. (ISCAS)*, Paris, France, May/Jun. 2010, pp. 393–396.

[35] H. H. Chen, Y.-H. Huang, P.-Y. Su, and T.-S. Ou, "Improving video coding quality by perceptual rate-distortion optimization," in *Proc. Int. Conf. Multimedia Expo (ICME)*, Singapore, Jul. 2010, pp. 1287–1292.

[36] 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.

[37] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Rate-SSIM optimization for video coding," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 833–836.

[38] 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.

[39] 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.

[40] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H. 264/AVC standard," *IEEE Trans. Circuit. Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[41] S. Hu, H. Wang, S. Kwong, T. Zhao, and C.-C. J. Kuo, "Rate control optimization for temporal-layer scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 8, pp. 1152–1162, Aug. 2011.

[42] T. Yang, C. Zhu, X. Fan, and Q. Peng, "Source distortion temporal propagation model for motion compensated video coding optimization," in *Proc. Int. Conf. Multimedia Expo (ICME)*, Melbourne, VIC, Australia, Jul. 2012, pp. 85–90.

[43] Y. Xu and C. Zhu, "End-to-end rate-distortion optimized description generation for H.264 multiple description video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 9, pp. 1523–1536, Sep. 2013.

[44] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Toronto, ON, Canada, Jul. 2006, pp. 1929–1932.

[45] S. Channappayya, A. C. Bovik, and J. R. W. Heathh, "Rate bounds on SSIM index of quantized images," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1624–1639, Sep. 2008.

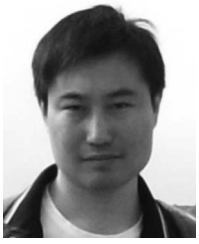
[46] I. Pao and M. Sun, "Modeling DCT coefficients for fast video encoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 4, pp. 608–616, Jun. 1999.

[47] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process. Image Commun.*, vol. 19, no. 2, pp. 121–132, Feb. 2004.

[48] *Subjective Video Quality Assessment Methods for Multimedia Applications*, ITU-T Recommendation P.910, 1999.

[49] G. Bjontegaard, *Calculation of Average PSNR Differences Between RD-Curves*, VCEG Document VCEG-M33, Apr. 2001.

[50] Y. H. Huang, T. S. Ou, P. Y. Su, and 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.
[51] M.-J. Chen and A. C. Bovik, "Fast structural similarity index algorithm," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Dallas, TX, USA, 2010, pp. 994–997.
[52] Z. Liu, Y. Sun, J. Zhang, X. Wen, and T. Su, "Video quality assessment based on fast structural similarity index algorithm," in *Proc. Int. Conf. Ubiquit. Future Netw. (ICUFN)*, Sapporo, Japan, 2012, pp. 336–339.
[53] H. B. Golestani and M. Ghanbari, "Minimisation of image watermarking side effects through subjective optimisation," *IET Image Process.*, vol. 7, no. 8, pp. 733–741, Nov. 2013.



Tiesong Zhao (S'08–M'12) received the B.S. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2006, and the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, in 2011. From 2011 to 2012, he was a Research Associate with the Department of Computer Science, City University of Hong Kong. He served as a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, ON, Canada, until 2013. He is currently working as a Research Scientist with the Department of Computer Science and Engineering, State University of New York at Buffalo, NY, USA. His research interests include image/video processing, visual quality assessment, and video coding and transmission.

Jiheng Wang (S'11) received the M.Math. degree in statistics computing from the University of Waterloo, Waterloo, ON, Canada, in 2011. He is currently pursuing the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada. He has been a Research Assistant with the Department of Electrical and Computer Engineering, University of Waterloo, since 2011. In 2013, he was with the Video Compression Research Group, Blackberry, Waterloo. From 2009 to 2010, he was a Research and Teaching Assistant with the Department of Statistics and Actuarial Science, University of Waterloo. His current research interests include 3-D image and video quality assessment, perceptual 2-D and 3-D video coding, statistical learning, and dimensionality reduction.

Zhou Wang (S'99–M'02–SM'12–F'14) received the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin in 2001. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image processing, coding, and quality assessment, computational vision and pattern analysis, multimedia communications, and biomedical signal processing. He has over 100 publications in the above areas with over 24 000 citations (Google Scholar). He was a recipient of the 2014 NSERC E. W. R. Steacie Memorial Fellowship Award, the 2013 IEEE Signal Processing Best Magazine Paper Award, the 2009 IEEE Signal Processing Society Best Paper Award, the 2009 Ontario Early Researcher Award, and the ICIP 2008 IBM Best Student Paper Award (as a senior author). He served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (2009)–(2014), *Pattern Recognition* since 2006, and the IEEE SIGNAL PROCESSING LETTERS (2006)–(2010) and a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (2013)–(2014) and (2007)–(2009), the *EURASIP Journal of Image and Video Processing* (2009)–(2010), and *Signal, Image and Video Processing* (2011)–(2013).

Chang Wen Chen (F'04) received the B.S. degree from the University of Science and Technology of China in 1983, the M.S.E.E. degree from the University of Southern California in 1986, and the Ph.D. degree from the University of Illinois at Urbana-Champaign in 1992. He is a Professor of Computer Science and Engineering with the University at Buffalo, State University of New York. He has been the Allen Henry Endow Chair Professor with the Florida Institute of Technology from 2003 to 2007. He was on the faculty of Electrical and Computer Engineering, University of Rochester from 1992 to 1996 and on the faculty of Electrical and Computer Engineering, University of Missouri-Columbia from 1996 to 2003. His research is supported by NSF, DARPA, Air Force, NASA, Whitaker Foundation, Microsoft, Intel, Kodak, Huawei, and Technicolor. He was a recipient of several research and professional achievement awards, including the Sigma Xi Excellence in Graduate Research Mentoring Award in 2003, the Alexander von Humboldt Research Award in 2009, and the State University of New York at Buffalo Exceptional Scholar—Sustained Achievement Award in 2012. He and his students were a recipient of the Best Paper Award and the Best Student Paper Award over the past two decades, for eight times. He has served as the Editor-in-Chief for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 2006 to 2009. He has been the Editor-in-Chief for the IEEE TRANSACTIONS ON MULTIMEDIA since 2014. He has been an Editor for several other major IEEE TRANSACTIONS AND JOURNALS, including the PROCEEDINGS OF IEEE, the IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, and the IEEE JOURNAL ON EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS. He has served as a Conference Chair for several major IEEE, ACM, and SPIE conferences related to multimedia, video communications, and signal processing. He is an SPIE Fellow.