Adaptive Downsampling Video Coding With Spatially Scalable Rate-Distortion Modeling

Ren-Jie Wang, Chih-Wei Huang, and Pao-Chi Chang, *Member, IEEE*

***Abstract*— Downsampling video coding, whereby downsampled frames are encoded, provides improved perceptual quality in rate-constrained situations. This method shows considerable advantages over other approaches, particularly in widespreading high-definition video formats. This paper provides a comprehensive analysis of downsampling video coding. The study proposes a spatially scalable rate-distortion (RD) model, comprising quantization-distortion and quantization-rate models, and develops an optimal encoding frame size determination framework. The proposed method achieves a gain up to 2.3 dB peak signal-to-noise ratio (PSNR) at 1 Mb/s when compared with conventional full frame size coding. The RD performance is close to the optimal scenario, in which the ideal frame size is obtained by heuristically performing downsampling coding in various allowable sizes.**

***Index Terms*— Downsampling, H.264, high definition (HD), rate distortion (RD) modeling, spatially scalable.**

I. Introduction

With advances in mobile communication and smart devices, the importance of high-definition (HD) video coding technology for the ever-growing use of wireless multimedia applications cannot be overemphasized. However, the available network bandwidth is not always adequate to stream such high-resolution video. In such circumstances, compression or transcoding with lower spatial resolution yields improved performance than coding with the original full-size video, because more bits are reserved per discrete cosine transform (DCT) coefficient [1]–[3]. Adaptive coding approaches that take advantage of spatial reduction can be classified into three categories: downsampling as mode selection, resolution transcoding, and downsampling as preprocessing.

In the case of downsampling as mode selection, a block scaling ratio is introduced as an encoding option and is integrated into the mode decision process. Nguyen *et al.* [4] proposed an adaptive downsampling mode decision process in the encoder. The modes that included various downsampling directions and block size ratios could be determined by analyzing residual block contents. To downsample the original block instead of the residual block, Choi *et al.* [5] proposed coding with various block size ratios with an adaptive motion vector (MV) prediction scheme. These methods enhance rate-distortion (RD) performance while sacrificing syntax conformation with video coding standards [6]. The modified bitstream cannot be successfully decoded by most off-the-shelf decoders.

The second category of frame size adaptation is resolution transcoding. In this approach, appropriate frame sizes as well as quantization parameters (QPs) for transcoding are selected to satisfy the bitrate constraint. In the method proposed by Shu and Chau [7], the bitrate of various resolution settings was estimated and the largest resolution satisfying the bitrate constraint to be encoded was selected. Yin *et al.* [8] modeled the impact of requantization, frame skipping, and spatial downsampling on RD in video transcoding. In the work proposed by Fling and Ro [9] and Jung *et al.* [10], the optimal combination of frame rate, size, and QP, based on RD models under the current constraints, was selected. However, these transcoding methods rely on full access of model parameters from the original encoded streams, which is not always the case. For example, in the situation of encoding raw frames in video recorder devices, the precoded information is not available.

The downsampling coding (DSC) approach preferred in this paper falls in the downsampling as the preprocessing category. In this approach, the original video frames are first downsampled and then encoded by a standard video codec. At the decoder side, the video sequence is decoded and then upsampled to the original resolution for displaying. The super-resolution technique of enhancing the quality of upsampled frames in the DSC decoder has been extensively studied. If a fixed downsampling ratio during encoding is known, the perceptual quality after super-resolution-assisted decoding can be improved by sending super-resolution-related side information [11] or by performing example-based training [12]. This paper holds that to achieve optimal performance, the downsampling ratio at the encoder has to be adaptively determined as well; dealing solely with the upsampling process might not achieve the optimum playback quality.

It has been empirically shown that, given the target bitrate, a corresponding frame size for the best encoded video quality exists [2], [3]. However, no automatic frame size determination scheme has been proposed. For still images, optimization of the downsampling ratio for JPEG image compression has been extensively studied [1]. However, the results cannot be directly applied to video coding, because the RD property of video coding is different from that of image coding. For video, Lee *et al.* [13] determined the frame size by maximizing the ratio between quality and computational complexity cost, while the RD model is empirically obtained with many parameters generated after the encoding stage. Rhee *et al.* [6] heuristically estimated the frame size based on the PSNR of

Manuscript received May 24, 2013; revised August 21, 2013 and October 16, 2013; accepted January 17, 2014. Date of publication January 28, 2014; date of current version October 29, 2014. This paper was recommended by Associate Editor J. Liang.

The authors are with the Department of Communication Engineering, National Central University, Jhongli 320, Taiwan (e-mail: rjwang@vaplab.ce.ncu.edu.tw; cwhuang@ce.ncu.edu.tw; pcchang@ce.ncu.edu.tw).

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/TCSVT.2014.2302519

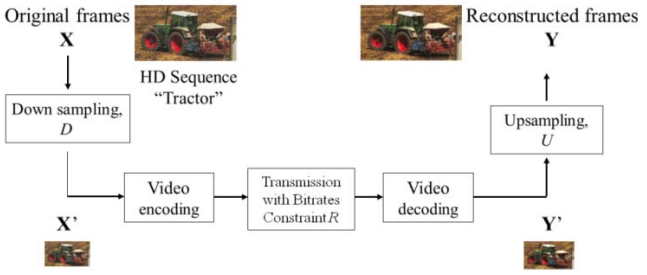


Fig. 1. Downsampling coding scheme. Original frames are downsampled before encoding and then upsampled to the original size for playback.

previously encoded frames on the frame size-distortion curve. However, if the frame size is estimated without RD modeling, the optimal frame size might not be obtained accordingly.

Empirical results in literatures show the existence of a sequence-dependent optimal scaling ratio that achieves the maximal quality for a given rate. It is a proper trade-off between encoding coding distortion and artifact from downsampling. This paper is motivated to investigate the impact of each component in DSC and develop a system to automatically find the optimal scaling ratio.

This paper proposes an adaptive DSC framework, with a comprehensive analysis of the spatially scalable RD model. The scalable RD of hybrid video coders was approximated by coupling downsampling distortion, quantization-distortion (QD), and quantization-rate (QR) models. A distortion-minimizing frame size searching scheme was also developed to determine the best encoding frame size for a given bitrate constraint. First, the distortion of the DSC was decomposed into two components, a downsampling component and a coding distortion component. The spatial scalability of each component was then analyzed to formulate the final RD model. The entire process depends solely on preencoding information, and is real-time implementable. As a result, the RD performance is superior to that of conventional full frame size or fixed downsampling ratio coding, and is close to the optimal scenario, in which the frame size is obtained by heuristically performing DSC in various allowable sizes.

The system architecture and analysis are described in Section II. The spatially scalable downsampling and coding distortion models are proposed in Section III. In Section IV, the adaptive frame size determination is addressed. Finally, the results and conclusion are presented in Sections V and VI, respectively.

II. Architecture Analysis of the Downsampling Coding System

The architecture of DSC is presented in Fig. 1. The original frame **X** was first downsampled to **X**' according to a scaling ratio s and then encoded with a QP that met the channel bitrate constraint R. At the decoder end, the video stream was decoded to **Y**' and then upsampled to **Y** in the original frame size. On the basis of the overall architecture, major enablers including distortion decomposition and the target problem of optimal frame size determination are introduced in this section. In this paper, the terms size and rate simply refer to the encoding frame size and bitrate.

A. Distortion Decomposition

To model the distortion with spatial scalability, the overall distortion was decomposed into a downsampling component and a coding component. Both components are dependent on a size parameter, which is the key to spatially scalable modeling. If X and Y are denoted as the input and output blocks of a DSC system in the size of M by M pels, then Y can be represented as

$$Y = U\left(\mathrm{DXD}^T + \delta_c\right)U^T \tag{1}$$

where $D(\cdot)D^T$ is the downsampling operation that shrinks the block from M by M to N by N; T denotes the matrix transpose; $U(\cdot)U^T$ is the upsampling operation that expands the block back to M by M; and δ_c represents the coding error with size N by N.

The difference δ between the input and the output of a DSC system can be derived from (1) as

$$\begin{aligned}\delta &= X - Y \\ &= \left(X - U\left(\mathrm{DXD}^T\right)U^T\right) - U\delta_c U^T \\ &\triangleq \delta_d - U\delta_c U^T. \end{aligned} \tag{2}$$

Equation (2) shows that δ can be decomposed into a downsampling error term δ_d and a coding error term $U\delta_c U^T$. The former is the difference between an original block and a block that has been through the downsampling and upsampling processes. The latter is the upsampled coding error of a downsampled block.

The average distortion of the DSC is defined as the mean square error (MSE) between the input and output frames of the DSC. By regarding input block X as a random process, the average distortion of one frame will be the expectation of the block distortion. Hence

$$\Delta_f = E\left\|\delta_d - U\delta_c U^T\right\|^2 \tag{3}$$

where the norm operation $\|.\|^2$ is defined as the mean of square for matrix elements. Because the inter-frame coding mode is usually chosen for general video sources, δ_c depends mainly on the inter-frame correlation. On the other hand, δ_d depends mainly on the variation of adjacent pels in a frame, representing intraframe correlation. It is thus assumed that δ_c and δ_d are uncorrelated. The actual correlation in video sequences is negligible, which will be shown in Section IV. Moreover, the upsampling procedure does not affect the mean of squares, i.e., the mean of squares for $U\delta_c U^T$ is identical to that of δ_c. The frame distortion becomes

$$\begin{aligned}\Delta_f &= E\left\|\delta_d\right\|^2 + E\left\|\delta_c\right\|^2 \\ &\triangleq \Delta_d + \Delta_c \end{aligned} \tag{4}$$

where Δ_d and Δ_c represent downsampling and coding distortion, respectively.

B. Optimal Encoding Frame Size

Based on the decomposed frame distortion in (4), each component can be modeled as a function of frame size and the QP. In particular, the downsampling distortion Δ_d depends on the scaling ratio $s = N/M$ and the characteristics of the current frame. The coding distortion Δ_c and the rate generated from the encoder depend on the scaling ratio, the characteristics of successive frames, and the QP. Therefore, the optimal frame size for minimizing the distortion can be obtained from the spatially scalable distortion and rate models.

Intuitively, for a given rate constraint and playback frame size, the downsampling distortion increases as the resolution decreases, whereas the coding distortion decreases with resolution shrinking because of the increasing rate per pel. The total distortion can be obtained by summing up the two parts of the distortions. Thus, an optimal scaling ratio can be obtained, which minimizes the overall distortion subject to the channel rate constraint R_c

$$\min_{s,QP} \Delta_f = \Delta_d(s) + \Delta_c(s, QP)$$

$$\text{s.t.} \quad R(s, QP) \le R_c. \tag{5}$$

III. Spatially Scalable Rate-Distortion Model

This section explores the impact of spatial scalability on RD modeling, taking into account the advantages of the previously mentioned distortion decomposition model. Scalable versions of the downsampling, QD, and QR models are proposed to complete the spatially scalable RD model.

A. Downsampling Distortion Modeling

The downsampling distortion was modeled in the transform domain for effective scaling over various downsampling ratios. This approach approximates the resizing results from commonly used bicubic methods. The downsampling distortion Δ_d for each size can be derived from the full-size transform coefficients instead of repeatedly performing expensive resizing processes in the spatial domain.

As mentioned, downsampling distortion is defined as the MSE between the original full-size block and the upsampled smaller block

$$\Delta_d = E\left\|\left(X - \text{UDXD}^T U^T\right)\right\|^2. \tag{6}$$

By the Parseval energy theorem, the average energy in the spatial domain is equal to that in the transform domain; hence

$$\Delta_d = E\left\|C\left(X - \text{UDXD}^T U^T\right)C^T\right\|^2 \tag{7}$$

where C represents the DCT transform matrix. The resizing approach is then used in the transform domain [14]. For downsampling, the smaller block can be constructed from the lower frequency coefficients of the original size. The downsampling operator D can be expressed as

$$D = \sqrt{\frac{N}{M}} C^B_{(NxN)} F_{(NxM)} C_{(MxM)} \tag{8}$$

where C and C^B are the forward and backward DCT transform matrices, respectively. Matrix F cuts the coefficients off after the index N. The average luminance is normalized by the ratio $\sqrt{N/M}$. For upsampling, zeros must be padded in the higher bands, and then a blurred block in the original size by backward transform with size M is obtained

$$U = \sqrt{\frac{M}{N}} C^B_{(MxM)} P_{(MxN)} C_{(NxN)} \tag{9}$$

where P pads M-N zeros to fill the size back to M. Substituting (8) and (9) into (7) yields

$$\Delta_d = E\left\|\text{CXC}^T - \text{PFCXC}^T F^T P^T\right\|^2. \tag{10}$$

The distortions for various sizes can thus be estimated from full-size transform coefficients. Downsampling distortion is the square of the sum of the coefficients in the truncated band. The downsampling distortion increases as the scaling ratio s decreases. Moreover, a video that has rich power in high-frequency bands suffers from large downsampling distortion.

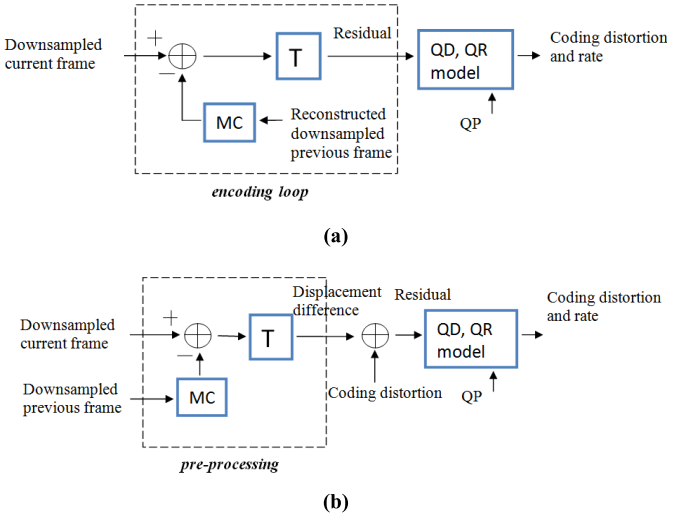


Fig. 2. Coding distortion and rate modeling. (a) In-loop model for rate control. (b) Out-loop estimation.

B. Coding Distortion Modeling

This section investigates the dependency of frame size on coding distortion and rate. In general, for the same contents, a shrunken video frame has less macroblock (MB) to be encoded, but the coding performance for each MB is diminished by a denser texture. In this paper, the size-dependent block distortion and rate were specifically modeled, so that the spatially scalable average distortion and total rate of a frame or a group of pictures (GOP) could be obtained.

Since the frame size has to be determined before entering the encoding loop, the RD estimation has to be performed out of the loop as well. An encoding loop begins with predetermined parameters including the frame size, so it requires repeating full encoding for many times to obtain RD performance across interested frame sizes. Therefore, our out-loop proposal, which estimates RD before actual encoding, is the key component to make the preprocessing possible. Fig. 2 shows the proposed out-loop approach. The QD and QR models for in-loop rate control take residual variance

and QP as inputs to generate estimated distortion and rate. Unfortunately, in the out-loop model, residual frames are not available yet. This paper obtains the required residual variance from the displacement difference and the previous coding distortion. Therefore, the modeling leads to another decomposition of the residual into the displacement difference and the coding distortion of previous frames. Finally, a key model of displacement difference for various scaling ratios is proposed. With size-scalable displacement difference modeling and out-loop QD and QR models, the spatially scalable out-loop QD and QR model for DSC is thus obtained.

Note that this paper proposes to practically keep the same resolution within a GOP as described in Section IV. By having the frame pair used for distortion analysis coming from the same GOP, the RD estimation becomes achievable. In the case of different resolutions between the current and the reference frames, the analysis of RD estimation becomes complicated.

1) General QD and QR Models: Almost all existing QD and QR modeling methods are performed in the transform domain because the transform coefficients for various video contents have similar distributions and the quantization in the codec is also operated in the transform domain. If the residual coefficient is assumed to be a random variable with a specific distribution, such as Laplacian [15] or Cauchy [16], the distortion can be modeled based on the variance of distribution and quantization stepsize q.

If it is assumed that the residual coefficients follow a Laplacian distribution, the probability density function (pdf) of the residual random variable is

$$p(y) = \frac{1}{\sqrt{2}\sigma_y} e^{-\sqrt{2}\cdot|y|/\sigma_y} \tag{11}$$

where σ_y is the standard deviation. Based on the quantization and reconstruction procedures in H.264/AVC [17], a closed form of the quantization distortion in MSE can be derived as follows [18]:

$$f(\sigma_y^2, q) = \sigma_y^2 - ((1-2\alpha)q + \sqrt{2}\sigma_y) \cdot \frac{q \cdot e^{-\sqrt{2}(1-\alpha)q/\sigma_y}}{1 - e^{-\sqrt{2}q/\sigma_y}} \tag{12}$$

where α is the length of the dead zone. According to the joint model (JM) reference software of H.264/AVC [19], α equals 1/3 for I-frames and 1/6 for P-frames and B-frames.

The entropy of the quantized residual coefficients can be obtained based on the pdf assumption. The entropy can also be applied to approximate the rate per pel statistically [20] and simplified to [21]

$$H(\sigma^2, q) = -p_0 \log p_0 - (1-p_0)\log\frac{c}{p} - \frac{2c\log p}{(1-p)^2} \tag{13}$$

where

$$p = \exp\left(-\frac{\sqrt{2}q}{\sigma}\right)$$

$$p_0 = 1 - \exp\left(-\frac{2\sqrt{2}\cdot q\cdot\alpha}{\sigma}\right)\sqrt{p}$$

$$c = \frac{1}{2}\exp\left(-\frac{2\sqrt{2}\cdot q\cdot\alpha}{\sigma}\right)\sqrt{p}(1-p).$$

In addition to the residual coefficients, the side information R_{side} (that includes the header, the MVs, and the coding modes), has to be encoded and transmitted within the bitstream and counted toward the overall bitrate. According to empirical observation, the rate per pel of side information can be approximately set as a constant, while the impact on the frame size selection is relatively low.

2) Residual Decomposition and Out-Loop Estimation: The residual can be decomposed into the displacement difference and the coding error, to reflect the distortion that corresponds to the video characteristics and the quantization, respectively [22]. The QD/QR estimation framework is built on this decomposition of the residual, to estimate the video quality before actual encoding in DSC.

For inter-frame prediction, a residual block r_k in the kth frame represents the difference between the current block X_k and the predicted block $Z\hat{X}_{k-1}Z^T$, which is a motion-compensated block in the previous reconstructed frame $\hat{X}_{k-1}$. Taking $ZX_{k-1}Z^T$, a motion-compensated block in the previous original frame, into the equation, the residual becomes

$$\begin{aligned} r_k &= X_k - Z\hat{X}_{k-1}Z^T \\ &= (X_k - ZX_{k-1}Z^T) + Z\left(X_{k-1} - \hat{X}_{k-1}\right)Z^T \\ &\triangleq \gamma + Z\delta_{c,k-1}M^T \end{aligned} \tag{14}$$

where γ is the displacement difference and $\delta_{c,k_{-1}}$ is the block coding error of the previous frame. It can thus be seen that the residual block is a performance mix of the current motion compensation and the encoding of the previous frame. By regarding a video sequence as a temporal stationary process, the variance of the residual becomes independent of the frame index k, and can be expressed as

$$\sigma_r^2 = \sigma_\gamma^2 + \Delta_c + 2\rho\sqrt{\sigma_\gamma^2}\sqrt{\Delta_c} \tag{15}$$

where σ_γ^2 and Δ_c are the variances of γ and $M\delta_c$, respectively. Because the displacement difference γ can be obtained from a premotion estimation (ME), σ_γ^2 can be calculated before the encoding procedures that include the prediction with rate-distortion optimization (RDO), transform, and quantization procedures. Finally, (15) is incorporated into (12) and (13), with the rate of side information included, to obtain

$$\Delta_c = f\left(\sigma_\gamma^2 + \Delta_c + 2\rho\sqrt{\sigma_\gamma^2}\sqrt{\Delta_c}, q\right) \tag{16}$$

$$R = \left(H\left(\sigma_\gamma^2 + \Delta_c + 2\rho\sqrt{\sigma_\gamma^2}\sqrt{\Delta_c}, q\right) + R_{\text{side}}\right)N_{\text{pel,full}} \times s^2 \tag{17}$$

where $N_{\text{pel,full}}$ is the number of pels in a full size frame. Distortion and rate functions are then built, which can be evaluated by taking the sequence-dependent σ_γ^2 and the quantization stepsize q as arguments. The coding distortion, which can be resolved by a root searching approach, is both the input and the output in (16).

3) Spatially Scalable Displacement Difference: Given the above out-loop RD model, the displacement difference is a function of the scaling ratio, i.e., the ME process must be performed when the scaling ratio is changed. Therefore, spatial scalability modeling for displacement difference is

highly desirable for enabling DSC applications, because the computational complexity can be greatly reduced. Generally, smaller frames have larger residual variation, which results in higher levels of distortion and higher rates per pel on average under the same QP [9]. In particular, the variance of γ is higher for downsampled frames because of the degraded ME performance. In addition to the original prediction errors in larger frames, downsampled frames suffer extra errors because of the relatively lower sampling resolution of the same objects. As a result, this paper proposes a scalable displacement difference model as follows.

A block X_k in a current original frame k can be represented as a shifted version of block X_{k-1} in a previous original frame $k-1$ plus the prediction error ε

$$X_k = S_v X_{k-1} S_h^T + \varepsilon \tag{18}$$

where S_v and S_h represent the vertical and horizontal shifts of corresponding blocks between two frames in full size resolution. Also, in the encoding process, motion compensation Z with sub-pel accuracy can be treated as a cascade of interpolation, integer shift, and sub-sampling [23], which is

$$Z = DS'U \tag{19}$$

where S' represents the estimated shift in the encoder; D and U are downsampling and upsampling matrices defined in Section II. The displacement difference γ for a specific downsampling ratio is represented as

$$\gamma = DX_k D^T - Z\left(DX_{k-1}D^T\right)Z^T. \tag{20}$$

Substituting (18) and (19) into (20), and assuming that the estimated shifts from the downsampled frames and the full frames are equal [23], i.e., $S' = S$, yields

$$\gamma = D\left(S_v\left(X_{k-1} - UDX_{k-1}D^TU^T\right)S_h^T + \varepsilon\right)D^T. \tag{21}$$

It is evident that the first major component ($X_{k-1} - \mathrm{UDX}_{k-1}D^TU^T$) is solely caused by down-then-up sampling. Considering the ME process in smaller frames, the MVs can be divided into two classes: either point-to-integer pels or point-to-interpolated pels. No interpolation error exists for MVs point-to-integer pels. However, for MVs point-to-interpolated pels, extra interpolation error exists and needs to be added to the displacement difference.

The proposed spatially scalable displacement difference model is thus represented as (22). When the MV for a smaller frame is an integer, that is, sampling on the same pels of an object in two frames, the displacement difference only results from ε. When motion for a smaller frame is at a sub-pel level, the impact of downsampling should be considered

$$\sigma_\gamma^2 = \begin{cases} E\left\|X - U\mathrm{DXD}^TU^T\right\|^2 + E\left\|\varepsilon\right\|^2, \\ \mathrm{mod}(MV_{v,\mathrm{full}}, 1/s) \neq 0 \text{ or } \mathrm{mod}(MV_{h,\mathrm{full}}, 1/s) \neq 0 \\ E\left\|\varepsilon\right\|^2, \quad \text{else.} \end{cases} \tag{22}$$

Because the downsampling distortion term decreases as the sampling rate decreases, (22) can reflect the variation of the displacement difference. By substituting (22) into (16) and (17), the spatial scalability is enabled. The following section shows the RD prediction by integrating these spatially scalable models in more detail.

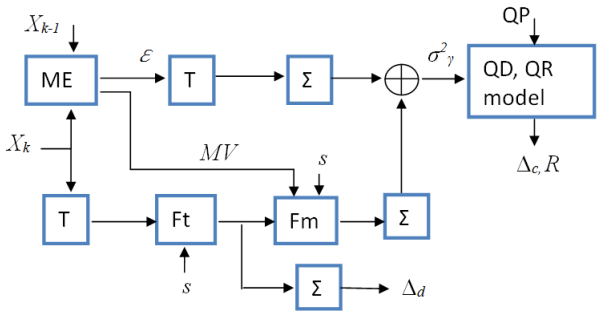


Fig. 3. Proposed estimation procedure for the RD at each size.

IV. Adaptive Frame Size Determination Algorithm

A. Overall Framework

The optimal encoding frame size was determined to minimize the distortion at a given rate, based on the spatially scalable RD model. The estimation procedure for the scalable RD model is illustrated in Fig. 3. With a particular scaling ratio s, the corresponding downsampling distortion Δ_d was calculated from the coefficients of a full-size frame, where a texture filter Ft represents the cutoff procedure of the frequency band in (10). The coding distortion Δ_c and the rate Rwere obtained from the out-loop RD model with the spatially scalable displacement difference model as proposed in Section III-B. The displacement differenceγ for each size consists of the imperfect prediction ε and the MV-dependent interpolation error, where a motion filter Fm represents the motion condition in (22). The operations were performed before the summations for each block. After all blocks were complete, the variance of each component in a frame was obtained, and then fed into the QD and QR models.

To provide an accurate displacement difference, a simplified ME and forward block-based transform was used to obtain the prediction error ε in the transform domain between two successive original frames. Furthermore, size-dependent operators were all located after the ME and transform procedures; that is, the ME and transform were performed only once in the original full frame size. With no need to perform complicated calculations for each feasible size, it was possible to evaluate the overall distortion repeatedly in real time.

For preME, UMHexagon fast ME from the H.264 reference software JM12.4 is applied. The search points are subsampled by hexagon masks with various scales. The maximal search range is restricted within ± 128 pels. Moreover, the partition size is set to 8×8 for the block-based ME. For the transform, frames are divided into 8×8 pel blocks; then the 2-D type-II DCT transform is applied on each block.

By feeding the estimated rate R, two distortions for each size, and QP into the object function (5), the optimal (s, QP) combination was obtained by searching in all feasible combinations. Specifically, downsampling distortion Δ_d for each size in the objective function (5) is obtained by the spatially scalable downsampling distortion model (10). For the

TABLE I
CORRELATIONS BETWEEN DOWNSAMPLING AND CODING ERROR IN VARIOUS QPS AND SCALING RATIOS

	s=1/4	1/3	1/2	3/4	1
QP=12	-0.020	-0.020	-0.025	-0.032	0
16	-0.015	-0.017	-0.022	-0.030	0
22	-0.004	-0.007	-0.019	-0.029	0
28	0.023	0.005	-0.014	-0.027	0
36	0.062	0.026	-0.003	-0.017	0
40	0.071	0.032	0.001	-0.017	0

TABLE II
CORRELATIONS BETWEEN DISPLACEMENT DIFFERENCE AND CODING ERROR IN VARIOUS FREQUENCY POSITIONS

	H0	H1	H2	H3	H4	H5	H6	H7
V0	0.105	0.248	0.367	0.303	0.105	0.248	0.367	0.302
V1	0.276	0.459	0.548	0.485	0.275	0.459	0.547	0.485
V2	0.406	0.560	0.611	0.573	0.406	0.559	0.611	0.573
V3	0.494	0.626	0.654	0.633	0.494	0.626	0.654	0.633
V4	0.106	0.249	0.368	0.303	0.106	0.249	0.367	0.303
V5	0.277	0.459	0.548	0.484	0.277	0.459	0.547	0.484
V6	0.409	0.560	0.611	0.573	0.409	0.560	0.611	0.573
V7	0.496	0.626	0.654	0.633	0.496	0.626	0.654	0.633

coding distortion Δ_c and rate R in the objective function (5), the out-loop QD function (16) and the QR function (17) are applied, where the displacement difference σ_γ^2 for each size is obtained by the spatially scalable displacement difference model (22).

B. Practical Approach and Parameter Selection

Practical issues, such as the update period and the parameter setting that enables the proposed adaptive downsampling algorithm to be performed on the existing video coding standards, need to be addressed. Taking compatibility with video coding standards and time efficiency into account, this paper proposes performing the downsampling in a GOP base. That is, the size is updated and kept fixed within a GOP (although the size could be different between two GOPs). In addition, the first two frames in the GOP are employed for estimation to reduce the computational overhead.

In the proposed framework, correlations between components cause approximation errors in the decomposition steps. Firstly, the correlation between downsampling and coding distortions was negligible, as shown empirically in Table I. The values were no larger than 0.071 across seven sequences in various scaling ratios and QPs. The correlations were obtained by heuristically coding results in various sizes. Exactly the same test settings and sequences listed in Table III are used to generate data in Tables I and II.

In contrast, the correlation between γ and δ_c cannot be ignored because both terms synchronously depend on inter-frame similarity. However, the variation among various sequences was not significant; the average difference among all testing sequences was 0.0635. The correlation for each frequency component was preset as shown in Table II, independent of QP and sequence. The set of values was obtained from the actual coding on pedestrian area in QP 28, which has minimal distance to the average of all test sequences. Due to the nature of transform and the fact that displacement difference and coding error are high-frequency signals, the correlation coefficients increase for the higher-frequency components. Correlation for DC also approaches zero as the analysis shown in [22]. The coefficient set has periodicity in both horizontal and vertical directions with intervals of 4 due to 4 × 4 transform in H.264 coding.

Moreover, the quasi-Newton method [24] was used for the implicit form of (16) to find the coding distortion. As discussed in Section III-B, the side information overhead was set as a constant, and the rate per pel for side information was set at 0.04, according to the average results obtained in real coding.

TABLE III
ENCODING CONFIGURATION IN VIDEO CODEC JM12.4

Resolution	Full HD (1920 × 1080)
Test sequences	Riverbed, Station, Rush-Hour, Sunflower, Pedestrian-area, Blue-sky, and Tractor
FramesToBeEncoded	250 frames
FrameRate	25 frames per second
RDO	On (high-complexity)
Fast motion estimation	UMHexagons
Number of reference frame	1 frame
Search range	±128 × scaling ratio *s*
Block mode	All on (16x16~4x4)
Sub-pel Motion Estimation	On (quarter level)
GOP	25 (structure: IPPP…)
Entropy coding	CABAC

V. EXPERIMENTAL RESULTS

The performance of the proposed spatially scalable RD model and adaptive DSC is demonstrated in this section. The proposed algorithm was implemented with H.264/AVC, and the performance on sample HD video sequences was assessed [25]. Our experiments used several self-coherent sequences with various properties to test the schemes over different contents. Real-time encoding scenarios with bandwidth variations, such as video conferencing and live video transmission, were the target situations; hence the low delay structure IPPP was selected. Both the RDO and the fast motion search algorithm UMHexagons were used, as listed in Table III.

The bicubic scheme was adapted for arbitrary frame size scaling. Given a scaling ratio, the interpolated position and the output pixel value were obtained by applying a 4 × 4 tap filter on neighboring pels. When reducing the size of an image, a sinc-like antialiasing filter corresponding to the scaling ratio was applied to limit the impact of aliasing on the downsampled image. The basis size M for transform and operation was

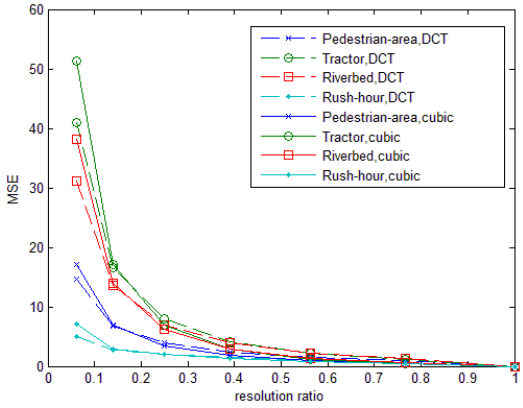


Fig. 4. Downsampling distortions for four sample sequences in DCT and bicubic methods, respectively.

set to 8, which provided the multiple scaling ratios commonly used at present. The available scaling ratios s were 2/8, 3/8, . . ., 7/8, and 1. In Sections V-A and V-B, the downsampling distortion, and coding QD and QR models are sequentially evaluated, and the overall RD performance of the proposed system is then shown.

A. Downsampling Distortion

In this section, the downsampling distortion in the test HD sequences is first shown and then compared with the adopted DCT domain modeling with the bicubic method used in general frame scaling. The downsampling distortions for various resolution ratios s^2, which is square of scaling ratio s, are shown in Fig. 4 and Table IV. Beginning with zero distortion at the full frame size (resolution ratio = 1), the downsampling MSE distortion increased as frame size decreased. In addition, the sequences with more power in the high spatial frequency bands had a higher level of distortion. For example, tractor, which contained a large grass area, suffered more distortion than the smoother pedestrian area. This result is in keeping with the analysis in Sections II and III.

The difference between the results from the bicubic and the DCT resizing methods is small, in the range from $s^2 = (1/2)^2$ to $(3/4)^2$. The Pearson correlation (PC) of the downsampling distortion between the DCT and bicubic methods when s^2 was in the range from $(3/8)^2$ to $(3/4)^2$ for all testing sequences was as high as 0.96. In applications that need to estimate distortions of various frame sizes, the reusability of transformation in the DCT approach enables low-complexity scalable modeling.

B. Spatially Scalable QD and QR Modeling

This section shows that the modeled rate and distortion behavior matches the actual encoder-generated cases for each QP and size. Fig. 5 shows the estimated and actual PSNR versus QP for each scaling ratio. Two sequences, tractor and rush hour, were selected as a combination to demonstrate outcomes from both high (31.66 dB at 2 Mb/s) and low (37.5 dB at 2 Mb/s) RD cost scenarios. The distortion at smaller sizes is higher than that for larger sizes at the same QP because more variations in displacement difference exist

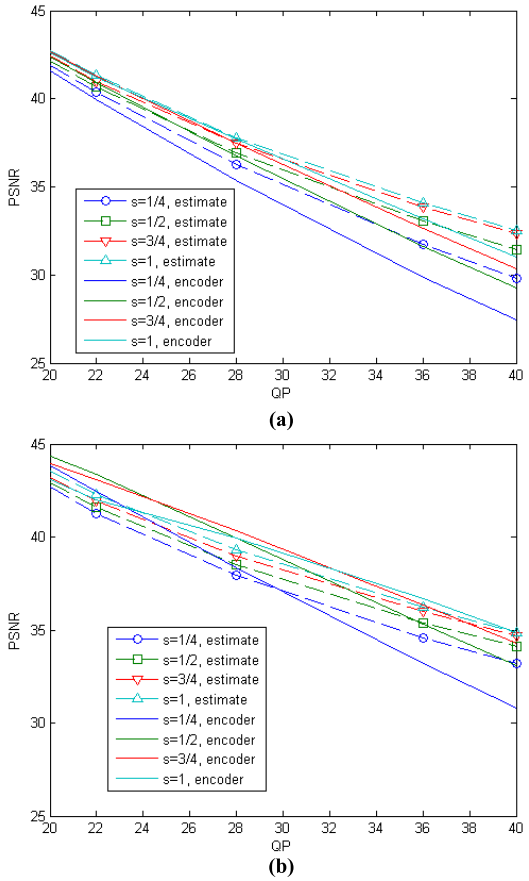


Fig. 5. Estimated and actual coding distortions against quantization parameter (QP). (a) Tractor (high RD cost). (b) Rush hour (low RD cost).

TABLE IV
DOWNSAMPLING DISTORTION RANGE AND THE DIFFERENCE BETWEEN DCT AND BICUBIC APPROACH FOR ALL TEST SEQUENCES

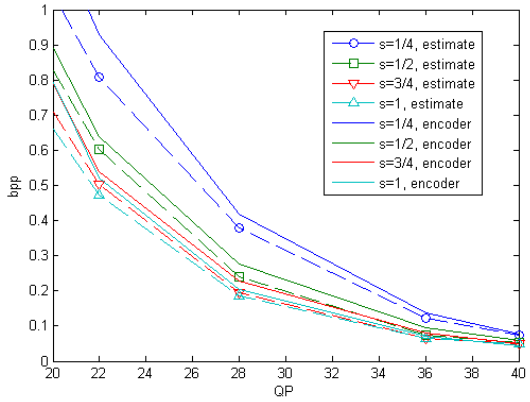
Sequence	$\Delta_{\pm}$ range (in s=[1/4 1])	MAD	MAD/range
Tractor	[41.0 0]	2.13	5.2%
Blue-sky	[40.8 0]	2.61	6.4%
Riverbed	[31.1 0]	1.55	5.0%
Station	[17.3 0]	0.80	4.6%
Pedestrian-area	[14.6 0]	0.64	4.4%
Rush-hour	[4.9 0]	0.47	9.6%
Sunflower	[4.3 0]	0.35	8.1%

at small sizes, as shown in Section III. The PCs between the modeled and the actual case are shown in Table V. For pedestrian area with a medium RD cost, the PC was 0.932 and the mean of the absolute difference (MAD) was 0.756. The QD estimation error between the estimate values and those generated by the encoder derived mostly from the distribution assumption, the out-loop estimation, and the pretraining fixed set of correlation values.

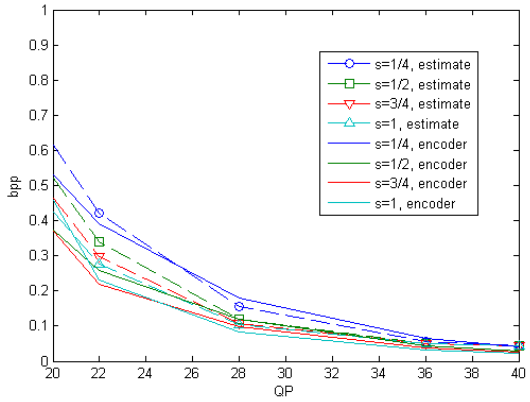
In terms of the spatially scalable QR model, the estimated and actual encoder output rates per pel for each size are shown in Fig. 6. The rate-size behavior was similar to the distortion-size analysis because both the rate and the distortion were estimated from the variation of the same residual signal.

TABLE V
PEARSON CORRELATIONS AND MADS BETWEEN THE PROPOSED MODEL AND THE ACTUAL RD IN VARIOUS SEQUENCES

	QD		QR	
Sequence	PC	Mad	PC	Mad
Pedestrian-area	0.932	0.756	0.927	0.040
Tractor	0.935	1.113	0.934	0.030
Riverbed	0.929	1.047	0.910	0.220
Rush-hour	0.923	0.878	0.926	0.027
Station	0.908	1.294	0.892	0.071
Sunflower	0.930	1.765	0.923	0.018
Blue-sky	0.921	1.467	0.930	0.028
Average	0.925	1.189	0.920	0.062



(a)



(b)

Fig. 6. Estimated and actual coding rates against quantization parameter. (a) Tractor (high RD cost). (b) Rush hour (low RD cost).

Smaller sizes yielded higher rates per pel, and the gap was obvious for small QP. Therefore, the behavior of the spatially scalable rate model can be obtained by the proposed method. The correlations and the MAD for the QR model are shown in Table V.

This paper also investigated the estimation error between the estimated and the actual results. In additional to the approximation error of the out-loop estimation, the estimation error for rate also came from the simplification of both the color component and the dependent coding [20]. However, the impact on the total rate of these components was not significant, because the luminance coefficient dominated the

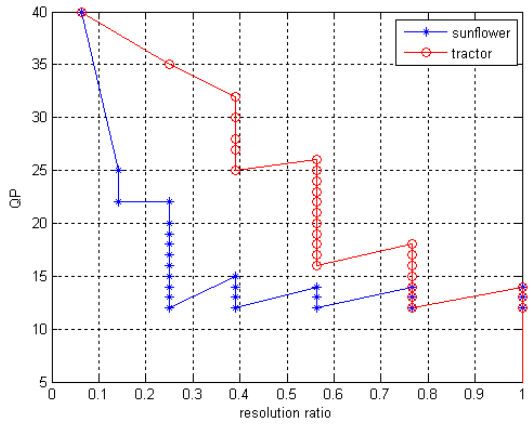


Fig. 7. Determined combination of resolution and QP for sunflower and tractor.

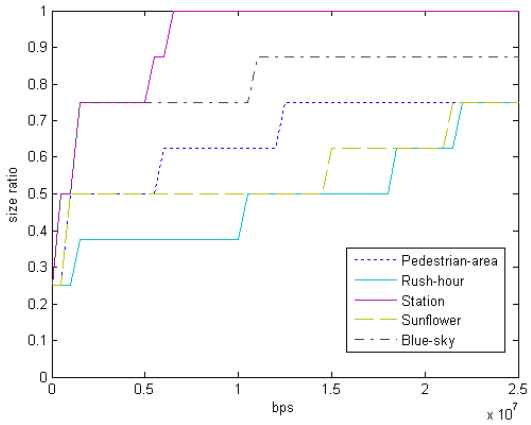


Fig. 8. Determined scaling ratio at each rate for five sample sequences.

total rate. The estimation error was limited to the range between QP = 22 and QP = 36, which is the commonly used QP range. The related rate and distortion variation of each size was sufficient to estimate the suitable frame size.

C. RD Performance With Adaptive Frame Size

The combination of QP and frame size selected by our scheme is shown in Fig. 7. It shows the path from the high bitrate with large size and small QP to the low bitrate with smaller frame and coarse QP. In addition to QP, the size is another parameter to control the rate. Both parameters can be used together to preserve quality for a reduced rate, considering that the content variation is helpful to understand the QP/size combination. As the rate decreases, sunflower prefers the coding with smaller size while maintaining QP since low downsampling distortion exists in the smooth contents. In contrast, tractor prefers the coding with lower QP while maintaining the size because of high downsampling distortion in the complicated structure in contents.

Fig. 8 shows the determined scaling ratio under various rate constraints. The scaling ratio was switched according to the available rate. In particular, the optimal size was decreased when the available rate was decreased, and the correct trade-offs between downsampling distortion and coding distortion

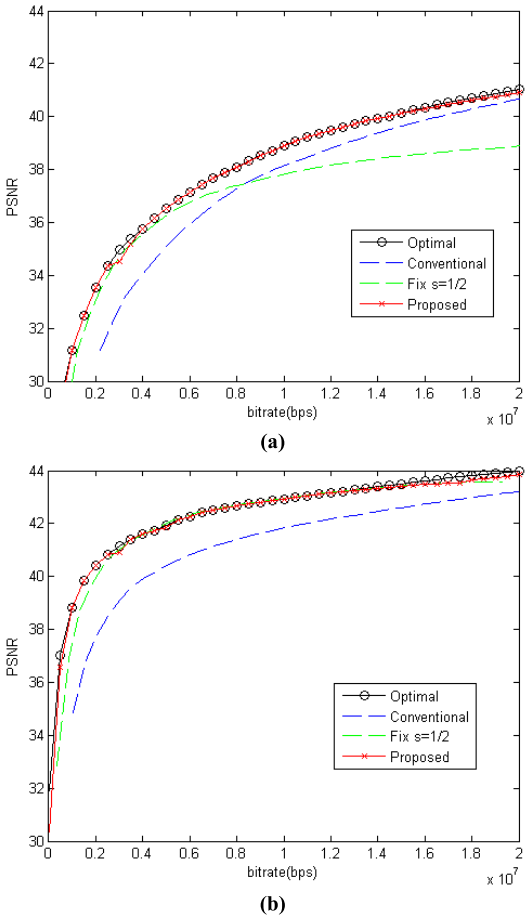


Fig. 9. Rate-distortion performance of the proposed method, conventional full frame size, and optimal case. (a) Tractor (high RD cost). (b) Rush hour (low RD cost).

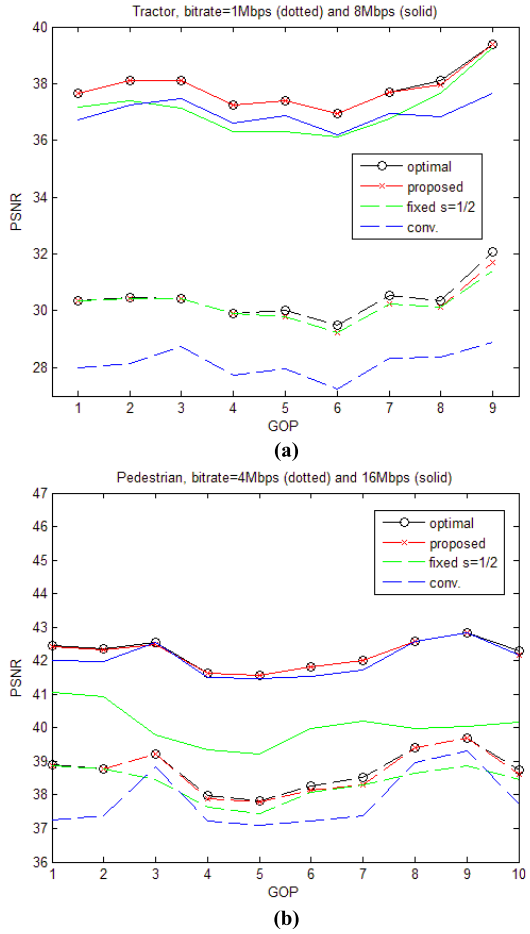


Fig. 10. PSNR comparison along time for various contents and bitrates. (a) Tractor. (b) Pedestrian area.

can be observed. Besides, in a longer video, scene changes with different RD properties happen all the time, so a sequence level determination may not fit all the scenes. If each testing sequence is considered as a scene, then Fig. 8 shows the case of different resulting scaling ratios along the time.

Fig. 9 compares the RD performance among the various methods: the proposed adaptive ratio method, the conventional coding method, downsampling with a fixed ratio ($s = 1/2$), and the optimal scenario. The conventional coding only adjusted the QP while maintaining the full frame size for encoding. The optimal frame size was obtained by exhaustively testing all available scaling ratios and QPs used in experiments. A combination of QP and scaling ratio with the highest PSNR subject to a rate constraint was selected.

It can be observed that the performance of the proposed adaptive DSC was comparable to the optimal scaling ratio. The conventional coding method performed poorly when the available rate was low, because too few bits per pel were available to reconstruct the frame well. DSC with a fixed ratio performed better at lower rates but suffered from serious downsampling distortion at higher rates.

The PSNRs' comparison along the time axis at different rates for two sequences is shown in Fig. 10. It can be observed that the proposed method consistently outperforms the fixed ratio and conventional methods along the time axis because of the content-adaptive scheme. Besides, the PSNR of fixed $s = 1/2$ becomes inconsistent with others at a high rate. It is because the downsampling distortion dominates in this situation.

Table VI shows the RD performance of all the test sequences in the various ranges of rates supported in present networks. It can be observed that the proposed method is effective for sequences with various texture and motion characteristics, such as the high-rate tractor and the low-rate station. On average, the proposed method achieved a 2.3-dB PSNR gain in the insufficient to medium rate (1 Mb/s), and a 1.72-dB gain over the fixed ratio method in the high-rate range (16 Mb/s). The proposed method always approaches the optimal scenario with a negligible PSNR drop.

When comparing to $s = 1/2$ at challenging lower rates, the adaptive nature leads us to advantages. Since the RD property is rather diverse across videos, the corresponding optimal encoding frame size is dynamic as well. By quick model-based assessment of couple frames, the suitable encoding sizes can be found. Therefore, it obtains improvements of 1.87 and 0.54 dB for riverbed and rush hour over $s = 1/2$ downsampling approach at a low rate of 1 Mb/s. The gain of 0.44 and 0.43 dB on average at 1 and 2 Mb/s over fixed downsampling is also quite positive.

TABLE VI
AVERAGE PSNR IMPROVEMENTS OF THE PROPOSED METHOD AT VARIOUS RATES (Mb/s) COMPARED WITH DIFFERENT METHODS

Sequence	Method	1Mbps	2M	4M	8M	16M
Pedestrian-area	Proposed	34.13	36.47	38.67	40.61	42.18
	△Conv.	1.69	1.36	0.83	0.49	0.15
	△s=1/2	0.11	-0.06	0.32	1.14	2.11
	△Opt.	-0.21	-0.10	-0.07	-0.05	-0.02
Tractor	Proposed	30.24	32.63	35.38	37.83	40.00
	△Conv.	2.08	1.97	1.54	0.88	0.44
	△s=1/2	0.03	-0.25	0.05	0.71	1.69
	△Opt.	-0.15	-0.30	-0.08	-0.02	-0.11
Riverbed	Proposed	27.01	28.47	31.00	33.31	36.34
	△Conv.	3.47	3.41	3.32	2.98	2.76
	△s=1/2	1.87	1.21	0.74	0.00	0.00
	△Opt.	0.00	-0.54	-0.17	-0.42	0.00
Rush-hour	Proposed	38.02	40.06	41.41	42.51	43.39
	△Conv.	3.22	2.56	1.61	1.21	0.77
	△s=1/2	0.54	0.25	-0.07	-0.08	0.00
	△Opt.	-0.03	0.00	-0.08	-0.08	-0.11
Station	Proposed	36.32	38.37	40.09	41.00	42.24
	△Conv.	1.21	0.23	0.38	0.00	0.00
	△s=1/2	0.20	0.76	1.55	1.75	2.47
	△Opt.	-0.04	-0.11	-0.01	-0.26	-0.12
Sunflower	Proposed	37.96	40.41	42.51	43.89	45.25
	△Conv.	3.18	1.99	1.17	0.51	0.44
	△s=1/2	0.14	-0.06	0.18	0.35	0.68
	△Opt.	-0.07	-0.07	-0.02	-0.13	-0.10
Blue-sky	Proposed	32.46	35.63	38.30	40.89	42.62
	△Conv.	1.30	0.54	0.10	0.06	-0.40
	△s=1/2	0.17	1.13	2.31	3.91	5.07
	△Opt.	-0.11	-0.03	-0.12	-0.02	-0.40
Average	Proposed	33.73	36.01	38.19	40.00	41.72
	△Conv.	2.31	1.72	1.28	0.88	0.59
	△s=1/2	0.44	0.43	0.73	1.11	1.72
	△Opt.	-0.09	-0.16	-0.08	-0.14	-0.12

TABLE VII
ACCURACY COMPARISON OF THE PROPOSED AND RHEE'S METHODS

△opt.	0.8Mbps	1.5Mbps	2Mbps
Proposed	-0.16dB	-0.24dB	-0.26dB
Rhee's	-0.35dB	-0.39dB	-0.46dB

At a high rate of 16 Mb/s, the optimal frame size was the full size for blue sky. Although the decision of the proposed method was also close to the full size, the estimation error sometimes resulted in decrease in performance. Favorably, the 0.4-dB PSNR loss at high 42.62-dB PSNR was almost imperceptible to the human eye.

The accuracy of the proposed method is compared with the approach proposed by Rhee *et al.* [6], as shown in Table VII.

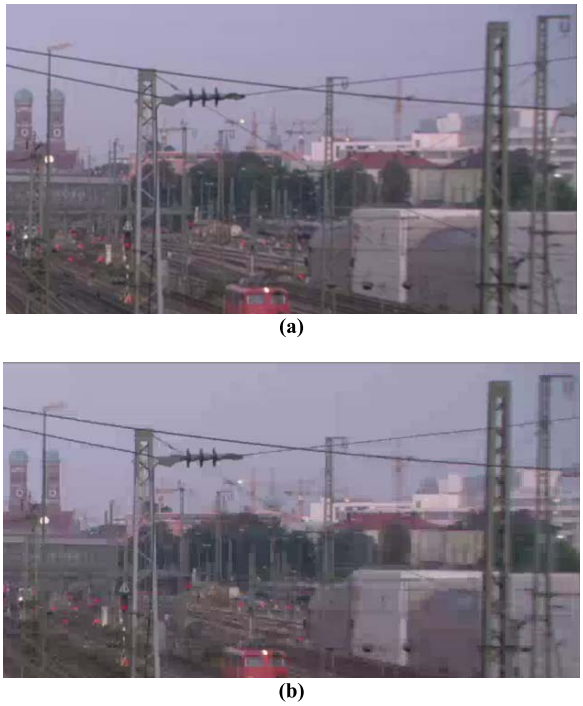


Fig. 11. Output sample of full HD sequence (1920 × 1080) station under the rate constraint 0.8 Mb/s. Only the up-right corner (960×540) of the frame is shown. (a) Adaptive downsampling. (b) Conventional full frame size coding.

Rhee's work heuristically estimated the frame size based on the PSNR of previously encoded frames to construct the frame size-distortion curve. It gradually changed the frame size GOP by GOP to reach the best size. The same experimental settings in the paper were used for comparisons: two full HD sequences life and speed bag; GOP size 30; and 150 frames. It can be observed that the proposed method performs closer to the optimal case than the related work does due to the direct model-based size decision. Moreover, the encoding frame size is expected to change more frequently using Rhee's method, which may result in more visual artifact.

Fig. 11 shows the output sample using the proposed and conventional coding under a rate constraint for station. Conventional coding results in an obvious blocking effect because of the block-based transform with a high QP. The frame reconstructed with DSC is smoother, that is, more comfortable to human vision.

The artifact introduced by the proposed adaptation is another issue. According to our observation, some strange artifacts can only be noticed when large scaling ratio variation happens during playback of coherent video contents. Favorably, the determined scaling ratio across coherent contents at a fixed data rate is relatively stable using our method. A large scaling ratio change mostly comes with content change, so the scaling artifact is less noticeable. Therefore, appropriately adjusting the coding frame size can still provide a satisfactory experience.

D. Computational Complexity

The encoding frame size searching process introduces relatively little extra complexity, while still achieving target

TABLE VIII
COMPUTATIONAL OVERHEAD FOR SIZE DECISION AND TOTAL ENCODING TIME OF THE PROPOSED METHOD. THE VALUE IS RELATIVE TO CONVENTIONAL FULL FRAME SIZE CODING

	Size decision overhead		Total encoding time (including overhead)		
Sequence	ME	Transform	1Mbps	4Mbps	16Mbps
Pedestrian-area	0.26%	1.24%	22.5%	23.2%	56.6%
Tractor	0.18%	0.92%	22.9%	37.2%	54.9%
Riverbed	0.26%	0.66%	6.40%	15.7%	27.4%
Rush-hour	0.22%	0.94%	6.63%	13.8%	24.8%
Station	0.16%	0.89%	19.9%	51.1%	100%
Sunflower	0.18%	0.99%	21.4%	22.7%	39.3%
Blue-sky	0.17%	1.00%	21.6%	48.5%	83.6%

results. Regarding the frequency of size changing, this paper suggests keeping the size fixed within a GOP, and so only the first two frames in the GOP are employed for estimation. Besides, owing to the spatially scalable model, the size determination preprocessing only needs to perform computationally heavy steps, such as ME and transform, once on the full-size frame. Moreover, if the ME results are reused for the first P frame coding after the size has been determined for speedup [26], the overall computation overhead can be further reduced.

Finally, because of processing in smaller frames, the overall encoding and decoding complexity benefits from DSC as well. Therefore, DSC is especially suitable for bandwidth- and computation-limited mobile scenarios.

Table VIII shows preprocessing overhead and encoding time relative to full frame size coding. A personal computer with i7 processer, 8 GB memory, and Windows 7 was used. It can be observed that the time overhead for frame size estimation is very limited compared with the conventional coding procedure. The overall time saving ratio depends on the rate and the contents. From Table VIII, the encoding time can be significantly saved because of small frame sizes. Note that the complexity of downsampling and upsampling is not shown because it is relatively simple, and moreover, downsampling and upsampling can be implemented in GPU, which can accelerate the process significantly.

The computational requirement for the optimum scheme is shown. The optimal size is determined by actual RD data that are only obtained after encoding all allowable frame sizes. The computational time is thus significantly higher than conventional full frame size coding. The relative complexity to full size coding of the optimal scheme is 254% if the available scaling ratio equals 2/8, 3/8, . . ., 8/8.

VI. CONCLUSION

In this paper, downsampling video coding was comprehensively explored. Spatially scalable downsampling distortion and coding distortion versus rates were modeled in the transform domain. Moreover, an effective size decision procedure was proposed under the DSC framework. The results show that the RD variation for each frame size can be closely approximated. The adaptive DSC provided a 2.3-dB PSNR improvement over the conventional full-size coding method at 1 Mb/s, and was close to the performance of the ideal scaling ratio.

In future research, the standard codec could be replaced by HEVC [27], and the RD analysis for high efficiency video coding (HEVC) may be different from H.264. Also, the spatially scalable RD model could be applied to resolution transcoding. This would provide improved performance, because more accurate parameters, such as MV and residual variance, could be obtained directly from the received bitstream. Moreover, the analysis framework proposed in this paper could also be applied in the field of optimal frame rate selection [28].

REFERENCES

[1] A. Bruckstein, M. Elad, and R. Kimmel, "Down scaling for better transform compression," *IEEE Trans. Image Process.*, vol. 12, no. 9, pp. 1132–1144, Sep. 2003.

[2] E. C. Reed and J. S. Lim, "Optimal multidimensional bit-rate control for video communication," *IEEE Trans. Image Process.*, vol. 11, no. 8, pp. 873–885, Aug. 2002.

[3] C. A. Segall, M. Elad, P. Milanfar, R. Webb, and C. Fogg, "Improved high-definition video by encoding at an intermediate resolution," *Proc. SPIE*, vol. 5308, pp. 1007–1018, Jan. 2004.

[4] V. A. Nguyen, Y. P. Tan, and W. S. Lin, "Adaptive downsampling/upsampling for better video compression at low bit rate," in *Proc. IEEE ISCAS*, May 2008, pp. 1624–1627.

[5] W. I. Choi, J. Yang, and B. Jeon, "Macroblock-level adaptive dynamic resolution conversion technique," *Proc. SPIE*, vol. 6391, pp. 639103-1–639103-8, Oct. 2006.

[6] C. E. Rhee, J. S. Kim, and H. J. Lee, "Bitrate control using a heuristic spatial resolution adjustment for a real-time H.264/AVC encoder," *EURASIP J. Adv. Signal Process.*, vol. 1, no. 1, pp. 1–12, Apr. 2012.

[7] H. Y. Shu and L. P. Chau, "The realization of arbitrary downsizing video transcoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 4, pp. 540–546, Apr. 2006.

[8] P. Yin, A. Vetro, and B. Liu, "Rate-distortion models for video transcoding," *Proc. SPIE*, vol. 5022, pp. 479–488, Jan. 2003.

[9] Y. J. Fling and Y. M. Ro, "Joint control for hybrid transcoding using multidimensional rate distortion modeling," in *Proc. ICIP*, vol. 4. Oct. 2004, pp. 2789–2792.

[10] Y. J. Jung, T. C. Thang, and Y. M. Ro, "Distortion measures in MPEG-compressed domain for multidimensional transcoding," in *Proc. IEEE 7th Workshop Multimedia Signal Process.*, Nov. 2005, pp. 1–4.

[11] D. Barreto, L. D. Alvarez, R. Molina, A. K. Katsagelos, and G. M. Callico, "Region-based super-resolution for compression," *Multidimensional Syst. Signal Process.*, vol. 18, nos. 2–3, pp. 59–81, Sep. 2007.

[12] M. Shen, P. Xue, and C. Wang, "Down-sampling based video coding using super-resolution technique," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 6, pp. 755–765, Jun. 2011.

[13] H. Lee, Y. Lee, J. Lee, D. Lee, and H. Shin, "Design of a mobile video streaming system using adaptive spatial resolution control," *IEEE Trans. Consum. Electron.*, vol. 55, no. 3, pp. 1682–1689, Aug. 2009.

[14] H. W. Park, Y. S. Park, and S. K. Oh, "L/M-fold image resizing in block-DCT domain using symmetric convolution," *IEEE Trans. Image Process.*, vol. 12, no. 9, pp. 1016–1034, Sep. 2003.

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

[16] N. Kamaci, Y. Altinbasak, and R. M. Mersereau, "Frame bit allocation for the H.264/AVC video coder via Cauchy density-based rate and distortion models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 994–1006, Aug. 2005.

[17] *ISO/IEC ITU-T Rec. H264: Advanced Video Coding for Generic Audiovisual Services*, ISO-IEC MPEG & ITU-T VCEG, Geneva, Switzerland, May 2003.

[18] D. S. Turaga, Y. Chen, and J. Caviedes, "No reference PSNR estimation for compressed pictures," *Signal Process. Image Commun.*, vol. 19, no. 2, pp. 173–184, Feb. 2004.

[19] *H.264/AVC Reference Software: Joint Model (JM)* [Online]. Available: http://iphome.hhi.de/suehring/tml/index.htm, accessed Aug. 2013.

[20] 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.
[21] C. H. Wu, Y. C. Tseng, and W. H. Peng, "Analytical mode-dependent rate and distortion models for H.264/SVC coarse grain scalability," in *Proc. IEEE ISCAS*, May 2012, pp. 1903–1906.
[22] L. Guo, O. C. Au, M. Ma, Z. Liang, and P. H. W. Wong, "A novel analytic quantization-distortion model for hybrid video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 5, pp. 627–641, May 2009.
[23] T. Wedi and H. G. Musmann, "Motion- and aliasing-compensated prediction for hybrid video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 577–586, Jan. 2003.
[24] C. G. Broyden, "The convergence of a class of double-rank minimization algorithms," *IMA J. Appl. Math.*, vol. 6, no. 3, pp. 222–231, 1970.
[25] *Standard HD Test Sequences* [Online]. Available: http://media.xiph.org/video/derf/ftp.ldv.e-technik.tu-muenchen.de/pub/, accessed Aug. 2013.
[26] R. Kumar and V. Patil, "An efficient motion vector composition scheme for arbitrary frame down-sampling video transcoder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 9, pp. 1148–1152, Sep. 2006.
[27] G. J. Sullivan, J. 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.
[28] M. C. Chien, R. J. Wang, C. H. Chiu, and P.C. Chang, "Quality driven frame rate optimization for rate constrained video encoding," *IEEE Trans. Broadcasting*, vol. 58, no. 2, pp. 200–208, Jun. 2012.

Ren-Jie Wang received the B.S. degree from National Quemoy University, Kinmen, Taiwan, and the M.S. degree from National Central University, Taoyuan, Taiwan, in 2007 and 2009, respectively. He is currently pursuing the Ph.D. degree at National Central University.

His research interests include video/image compression, scalable coding, video surveillance, and image/video retrieval.

Chih-Wei Huang received the B.S. degree from National Taiwan University, Taipei, Taiwan; the M.S. degree from Columbia University, New York, NY, USA; and the Ph.D. degree from University of Washington, Seattle, WA, USA, in 2001, 2004, and 2009, respectively, all in electrical engineering.

He joined the Department of Communication Engineering, National Central University, Taoyuan, Taiwan, in 2010. He is currently an Assistant Professor with the Information Processing and Communications Laboratory. From 2006 to 2009 he was an Intern Researcher with Siemens Corporate Research and Microsoft Research. He has published in the areas of wireless networking, multimedia communications, digital signal processing, and information retrieval.

Pao-Chi Chang (M'86) received the B.S. and M.S. degrees from National Chiao-Tung University, Hsinchu, Taiwan, and the Ph.D. degree from Stanford University, Stanford, CA, USA, in 1977, 1979, and 1986, respectively, all in electrical engineering.

He was a Research Staff Member with the Department of Communications, IBM T. J. Watson Research Center, New York, NY, USA, from 1986 to 1993. At Watson, his work centered on high-speed switching systems, efficient network design algorithms, and multimedia conferencing. In 1993 he joined the faculty of National Central University, Taoyuan, Taiwan, where he is currently a Professor with the Department of Communication Engineering and the Department of Electrical Engineering, and the Supervisor of the Video-Audio Processing Laboratory. He was a Visiting Professor with Stanford University from 2000 and 2004. His research interests include speech/audio coding; video/image compression; digital watermarking and data hiding, multimedia retrieval, and multimedia delivery over packet and wireless networks.