An Adaptive Control for the Propagation Errors Incurred by DCT Coefficient-Dropping Transcoder

Jin-soo Kim, Jae-gon Kim, Kwang-deok Seo, and Mong-han Yun

This paper presents a new distortion control scheme with a simple estimation model for the propagation errors incurred by dropping some parts of the bitstream in a frame dropping–coefficient dropping (FD-CD) transcoder. The primary goal of this paper is to facilitate bit-rate conversions and rate-distortion controls in the compressed domain without introducing a full decoding and re-encoding system in the pixel domain. First, the error propagation behavior over several frame sequences due to coefficient dropping is investigated on the basis of statistical and empirical properties. Then, such properties are used to develop a simple estimation model for the CD distortion accounting for the characteristics of the underlying coded-frame. Finally, the proposed estimation model allows us to determine the amount of coefficient dropping and to effectively allocate rate-distortions into coded-frames. Experimental results show that the proposed estimation model accurately describes the characteristics of propagation errors adaptively in the compressed domain and can be easily applied to distortion control over different kinds of video sequences.

Keywords: Coefficient dropping, propagation error, distortion control.

I. Introduction

In the framework of universal multimedia access (UMA), one challenge for video transmission is to deliver video content through heterogeneous network channels matching the diversity of client devices [1]. As one approach, video transcoding schemes are commonly used [2]. In particular, transcoding based on dropping some parts of the bitstream in the compressed domain is used mainly due to its low computational complexity and simple implementation [3], [4]. Since a frame memory is not required, and there is no need for an inverse discrete cosine transform (IDCT), regardless of the techniques used to achieve the reduced bit-rate, these schemes are relatively simple. This paper focuses on the frame dropping–coefficient dropping (FD-CD) scheme, which provides trade-offs between spatial and temporal qualities and extends the range of rate reduction. By combining frame-dropping and DCT CD, it is possible to simply adapt the bit rate of a pre-coded video to the available dynamic bandwidth, especially in streaming applications. Furthermore, the FD-CD scheme can achieve rate shaping across a wide range of rate reduction while maintaining spatial-temporal quality as much as possible. For a particular FD-CD operation, in which non-reference frames (namely, B frames) are dropped to avoid the decoding dependency, CD is applied to meet the available bandwidth quite accurately by adjusting the number of dropped coefficients.

Some previous studies on FD-CD transcoders have focused on the optimal selection of the number of dropped coefficients within a frame [1], [5]-[9]. The dynamic rate shaping (DRS) approach was introduced as a technique to adapt the rate of compressed video bit-streams [10]. The behavior of DRS was analyzed for both constrained and unconstrained cases by

---

Manuscript received Jan. 10, 2007; revised Aug. 14, 2007.

Jin-soo Kim (phone: + 82 42 821 1147, email: jskim67@hanbat.ac.kr) and Mong-han Yun (email: monghan@pinetelecom.com) are with the Department of Multimedia Engineering, Hanbat National University, Daejeon, Rep. of Korea.

Jae-gon Kim (email: jgkim@kau.ac.kr) is with the School of Electronics, Telecommunications and Computer Engineering, Korea Aerospace University, Gyeonggi-do, Rep. of Korea.

Kwang-deok Seo (email: kdseo@yonsei.ac.kr) is with the Department of Computer and Telecommunications Engineering, Yonsei University, Gangwon-do, Rep. of Korea.

assuming a first order autoregressive source [9]. In [7] and [8], by incorporating frame dependency into the rate shaping process and using feedback information, the shaping performance could be improved. In these FD-CD schemes, each coded DCT coefficient is treated as an intra-coded one, so its propagated/accumulated errors are ignored [1], [10], [11]. However, these errors are also accumulated, which often results in significant visual quality degradation. In this paper, we analyze the characteristics of the propagated errors incurred by CD and propose an effective estimation model which adaptively and efficiently describes the characteristics of propagation errors in the compressed domain. Furthermore, based on this model, a method is developed to determine the number of coefficients to drop to effectively allocate rate-distortions into coded frames.

II. Overview of FD-CD Transcoding Schemes

The FD method is a temporal down-sampling technique to achieve large scale bit rate changes by dropping some frames. To reduce the complexity of re-encoding, we adopt the more straightforward operations of an FD that drops B and/or P frames which do not have decoding dependency in the unit of group of pictures (GOP) by taking into account the GOP structure.

Unlike FD, CD provides the ability to meet the available bandwidth quite accurately by adjusting the number of dropped coefficients. Our CD work in this paper is based on a previous work on dynamic rate shaping (DRS) [10], as shown in Fig. 1. We assume that a set of high frequency DCT coefficient bits run-length coded at the end of each block is eliminated as in the constrained DRS case in [9] and [10]. In another unconstrained DRS [9] and [10], the dropping of an arbitrary set of coefficients is considered and done by optimization search. Because the zigzag scanning pattern of a DCT block provides a quite successful ordering of the DCT coefficients according to their importance, the sub-optimum result usually achieved by the constrained DRS varies very slightly from the unconstrained scheme in terms of decoded video quality [10]. Therefore, we drop AC DCT coefficients to avoid somewhat complicated syntax changes and to maintain the minimum necessary quality.

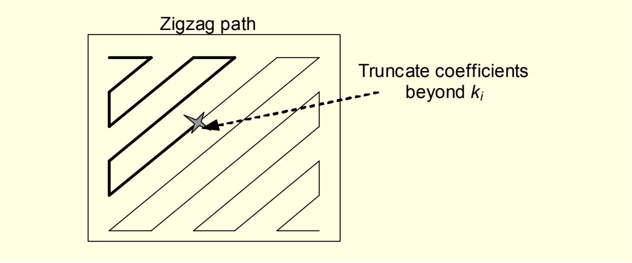


Fig. 1. CD operation of the constrained DRS scheme.

There are several CD control algorithms [1], [6]-[9], [12], [13]. In [5], a two-stage R-D optimization approach was proposed to solve the error concealment aware rate shaping (ECARS) which used the model-based hyper-surface and the hill climbing based refinement. In [1], [12], [13] two CD methods were presented: the uniform rate-based CD (URCD) and the Lagrange optimization CD (LOCD). Using URCD, a uniform ratio of bits is truncated from each frame and the number of truncated bits is uniformly distributed into each DCT block. This uniform distribution is simple but does not yield even distortion to every frame. Using LOCD, for the *j*-th frame with bit budget B_{keep}, the problem is to find a new breakpoint set $\vec{k}=(k_1,k_2,\cdots,k_M)$, where M is the total number of blocks per frame. The problem is formulated as

$$\underset{\vec{k}=(k_1,k_2,\ldots,k_M)}{\text{minimize}}\{\sum_{i=1}^{M}(\sum_{k=k_i+1}^{N-1} {z_{i,k}}^2+\lambda\sum_{k=0}^{k_i} b_{coeff}^{i,k})\}, \qquad (1)$$

where $z_{i,k}$ and $b_{coeff}^{i,k}$ are the dequantized-DCT coefficient and the amount of coded bits for the *k*-th symbol in the *i*-th block ($i=1, 2,\cdots, M$), respectively; and N is the number of coefficients. This problem is to find a suitable λ, such that when unconstraint problem (1) reaches its optimum, we have $\left|B_{Keep}-\lambda\sum_{i=1}^{M}\sum_{k=0}^{k_i} b_{coeff}^{i,k}\right|\le\varepsilon,\ \varepsilon>0$. This is done based on an bisectional iteration approach. In each iteration, a Lagrange multiplier λ is used to find a corresponding optimum solution, that is, the truncation point set $\vec{k}(\lambda)=(k_1,k_2,\cdots,k_M)_\lambda$. This is implemented using the exhausted search, which is not a heavy load. Once $\left|B_{Keep}-\lambda\sum_{i=1}^{M}\sum_{k=0}^{k_i} b_{coeff}^{i,k}\right|\le\varepsilon,\ \varepsilon>0$ is satisfied, the iteration stops. This LOCD ensures that the given bit budget is distributed so that all blocks have the same incremental improvement in quality for the same increment in bit rate.

III. Statistical Property and Estimation Model for the Propagation Errors Incurred by Coefficient-Dropping

1. Statistical Properties of CD Errors

As shown in Fig. 1, we assume that only a contiguous string of DCT coefficients at the end of each block is dropped (constrained DRS) [9]. Let $y_j=\{y_{j,i}\}_{i=1,\cdots,M}$ denote the frame *j* having *M* blocks, where $y_{j,i}$ is the decoded block *i* of frame *j*. Based on these notations, let $\hat{y}_{j-1}$ and $\hat{y}_j$ represent the decoded frames *j*-1 and *j*, respectively, degraded by CD operations. Then, decoded blocks with and without CD can be expressed as

$$\begin{aligned} y_{j,i} &= C_{j,i}(y_{j-1}) + e_{j,i}, \\ \hat{y}_{j,i} &= C_{j,i}(\hat{y}_{j-1}) + \hat{e}_{j,i}, \end{aligned} \tag{2}$$

where $c_{j,i}(y_{j-1})$ is the motion compensated (MC) component with reference to frame j-1, and $e_{j,i}$ is the coded residual component. Additionally, when the coefficients of which the scan order is greater than k_i are dropped, $\hat{e}_{j,i}$ is obtained from $e_{j,i}$ as shown in Fig. 1. Thus, the mean square error (MSE) distortion of block i of frame j, $D_{j,i}(k_i)$, is given by

$$\begin{aligned} D_{j,i}(k_i) &= \frac{1}{N}\left\| y_{j,i} - \hat{y}_{j,i} \right\|^2 \\ &= \frac{1}{N}\left\| C_{j,i}(y_{j-1,i}) - C_{j,i}(\hat{y}_{j-1,i}) + e_{j,i} - \hat{e}_{j,i} \right\|^2, \end{aligned} \tag{3}$$

where N is the block size measured by the number of pixels. Here, $C_{j,i}(y_{j-1,i}) - C_{j,i}(\hat{y}_{j-1,i})$ is the propagated error from the previous degraded frame j-1 by the MC operator and simply denoted as $a_{j,i}$, while $e_{j,i} - \hat{e}_{j,i}$ is the current CD error for block i of frame j.

Let $A_{j,i}(k)$, $E_{j,i}(k)$, and $\hat{E}_{j,i}(k)$ denote the k-th DCT coefficient of $a_{j,i}$, $e_{j,i}$, and $\hat{e}_{j,i}$ in the DCT-domain representation as

$$\begin{aligned} A_{j,i}(k) &= DCT_k(a_{j,i}), \\ E_{j,i}(k) &= DCT_k(e_{j,i}), \quad k = 0, \cdots, N-1, \\ \hat{E}_{j,i}(k) &= DCT_k(\hat{e}_{j,i}). \end{aligned} \tag{4}$$

Then, since coefficients greater than k_i along the scan order are dropped, based on the energy reservation property and the additive property of DCT [14], (3) can be expressed as

$$\begin{aligned} D_{j,i}(k_i) &= \frac{1}{N}\sum_{k=0}^{N-1}\left\{A_{j,i}(k) + E_{j,i}(k) - \hat{E}_{j,i}(k)\right\}^2 \\ &= \frac{1}{N}\sum_{k=0}^{N-1} A^2{}_{j,i}(k) + \frac{2}{N}\sum_{k=0}^{N-1} A_{j,i}(k)(E_{j,i}(k) - \hat{E}_{j,i}(k)) \\ &\quad + \frac{1}{N}\sum_{k=0}^{N-1}(E_{j,i}(k) - \hat{E}_{j,i}(k))^2 \\ &= \frac{1}{N}\sum_{k=0}^{N-1} A^2{}_{j,i}(k) + \frac{2}{N}\sum_{k=k_i+1}^{N-1} A_{j,i}(k)E_{j,i}(k) \\ &\quad + \frac{1}{N}\sum_{k=k_i+1}^{N-1} E^2{}_{j,i}(k). \end{aligned} \tag{5}$$

From this analysis, it is shown that the CD distortion involves not only the propagated errors (first term) and the current error (third term), but the correlated error (second term) as well. The average distortion over all blocks in the j-th frame, D_j, is given by

$$D_j = \frac{1}{M}\sum_{i=1}^{M} D_{j,i}(k_i). \tag{6}$$

Property 1. The MSE in the j-th frame can be approximated to $D_j \cong \frac{1}{NM}\sum_{i=1}^{M}\left\{\sum_{k=0}^{N-1} A^2{}_{j,i}(k) + \sum_{k=k_i+1}^{N-1} E^2{}_{j,i}(k)\right\}$. That is, the correlated error is negligible; thus, the distortion of the j-th frame is determined by the sum of the current error and the propagated error.

Proof. The proof is straightforwardly based on the fact that $A_{j,i}(k)$ and $E_{j,i}(k)$ are statistically independent, and their expected values are zero for NM>>1 [13], [14]. □

Furthermore, when CD can be applied to two consecutive frames, namely, frames (j-1) and j, (2) is expanded as

$$\begin{aligned} \hat{y}_{j+1,i} &= C_{j+1,i}(\hat{y}_j) + \hat{e}_{j+1,i} \\ &= C_{j+1,i}(C_{j,i}(\hat{y}_{j-1}) + \hat{e}_{j,i}) + \hat{e}_{j+1,i}. \end{aligned} \tag{7}$$

Generally, since universal MC algorithms deal with translation motion, we assume that the MC operator $C(\cdot)$is a linear function [15]. Then, the expression of (7) can be easily extended to the (j+1)th frame as

$$\hat{y}_{j+1,i} = C_{j+1,i}\left(\left\{C_{j,l}(\hat{y}_{j-1})\right\}_{l=1,\cdots,M}\right) + C_{j+1,i}\left(\left\{\hat{e}_{j,l}\right\}_{l=1,\cdots,M}\right) + \hat{e}_{j+1,i}. \tag{8}$$

Similarly to (4), DCT representations of each component for (8) are expressed as

$$\begin{aligned} G_{j+1,i}(k) &= DCT_k\left(C_{j+1,i}\left(\left\{C_{j,l}(y_{j-1} - \hat{y}_{j-1})\right\}_{l=1,\cdots,M}\right)\right), \\ A_{j+1,i}(k) &=, DCT_k\left(C_{j+1,i}\left(\left\{e_{j,l} - \hat{e}_{j,l}\right\}_{l=1,\cdots,M}\right)\right), \\ E_{j+1,i}(k) &= DCT_k(e_{j+1,i} - \hat{e}_{j+1,i}). \end{aligned} \tag{9}$$

By using these notations, the overall distortion of the (j +1)th frame can be summarized as property 2.

Property 2. When CD is applied to two consecutive frames, namely, frame (j -1) and frame j, the overall distortion of the (j +1)th frame, D_{j+1}, can be approximated to

$$D_{j+1} \cong \frac{1}{NM}\sum_{i=1}^{M}\left\{\sum_{k=0}^{N-1} G_{j+1,i}^2(k) + \sum_{k=0}^{N-1} A_{j+1,i}^2(k) + \sum_{k=k_i+1}^{N-1} E_{j+1,i}^2(k)\right\},$$

which means the overall distortions of the (j +1)th frame can be expressed by the sum of the current error and the propagated errors from the previous frames (j -1) and j.

Proof. The proof is similar to that of property 1.

$$D_{j+1,i}(k_i)=\frac{1}{N}\left\|y_{j+1,i}-\hat{y}_{j+1,i}\right\|^2$$

$$\begin{aligned}D_{j+1}&=\frac{1}{M}\sum_{i=1}^{M}D_{j+1,i}\left(k_i\right)\\&=\frac{1}{MN}\sum_{i=1}^{M}\Big\|C_{j+1,i}\left(\left\{C_{j,l}\left(y_{j-1}\right)\right\}_{l=1,..,M}\right)+C_{j+1,i}\left(\left\{e_{j,l}\right\}_{l=1,..,M}\right)\\&\qquad+e_{j+1,i}-C_{j+1,i}\left(\left\{C_{j,l}\left(\hat{y}_{j-1}\right)\right\}_{l=1,..,M}\right)\\&\qquad-C_{j+1,i}\left(\left\{\hat{e}_{j,l}\right\}_{l=1,..,M}\right)-\hat{e}_{j+1,i}\Big\|^2.\end{aligned}$$

Similarly to (5), based on the energy reservation property and the additive property of DCT [14], this is computed in the DCT domain. Moreover, since $G_{j+1,i}(k)$, $A_{j+1,i}(k)$ and $E_{j+1,i}(k)$ are independently determined and their expected values are zero for $NM>>1$, D_{j+1} can be simply approximated. □

Extending property 2 to frame (j+2), frame (j+3) and so on, the overall distortion in a frame can be expressed as the sum of the current error and the propagated errors from all previous frames.

2. Simulations for the Proof of Properties 1 and 2

For simulations, well-known MPEG test sequences with different classes (Class A: Akiyo, Mother and Daughter; Class B: Foreman, Coast Guard; Class C: Stefan, Table Tennis) are used. Table 1 summarizes the coded characteristics and classes of the test video sequences used in this study. (Note that this paper uses renumbering in all figures after removal of the B frames.)

First, the Foreman sequence in CIF format is coded at 1.5 Mbps with a GOP size = 15 and a sub-GOP size = 3. The sequence is transcoded to 800 kbps by CD with uniform rate reduction among frames and FD of all B-frames dropping. Figure 2 compares each component of the CD distortion (measured in MSE). As denoted in (5), "current error" is the DCT coefficient-drop distortion per frame and "CD distortion" represents the overall distortion including the errors propagated and measured in the decoded frames. In this simulation, to obtain "accumulated error," CD operation is applied to each coded frame independently. That is, CD operation is applied to all I-frames and not applied to the other frames. By decoding the transcoded bit stream, the accumulated error at the first P-frame is found. Then, CD operation is applied to all I-frames and the first P-frame, simultaneously, but it is not applied to the other frames. By decoding the transcoded bit stream, the accumulated error at second P-frame is obtained. The accumulated error of the other P-frames is found by using this step. On the other hand, the correlated error is found by subtracting the current error and the accumulated error from the CD distortion frame by frame. This result ensures that the correlated error term can be ignored while the current and the propagated errors are dominant. Therefore, property 1 is very useful.

Table 1. Test video sequences (compressed by MPEG-4 SP with GOP=15, Sub-GOP=3).

Test seq.	SC	Bit rate	N-intra	N-inter	N-not
Akiyo	A	1.2 Mb/s	0	1584	695.3
Mother & Daughter	A	800 kb/s	1.2	1582.8	673.4
Foreman	B	1.5 Mb/s	34.1	1549.9	224.6
Coast Guard	B	2 Mb/s	0.7	1583.3	84.1
Stefan	C	1.5 Mb/s	19.6	1564.4	350.7
Table Tennis	C	3 Mb/s	12.1	1571.9	185.8

SC: MPEG-4 sequence class.
Bit rate: bit rate before transcoding.
N-intra: the average number of intra-coded blocks per frame (PVOP)
N-inter: the average number of inter-coded blocks per frame (PVOP)
N-not: the average number of not-coded blocks in inter-coded MBs per frame (PVOP)

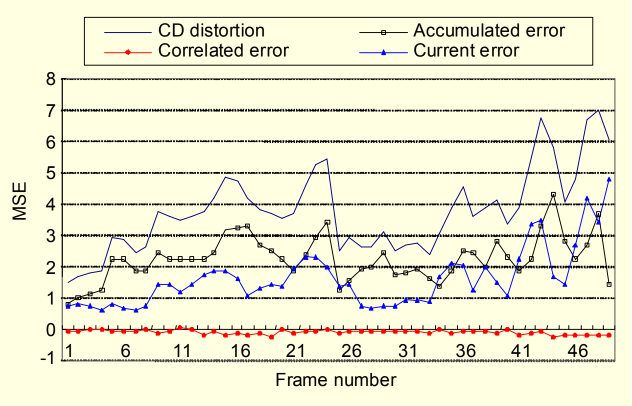
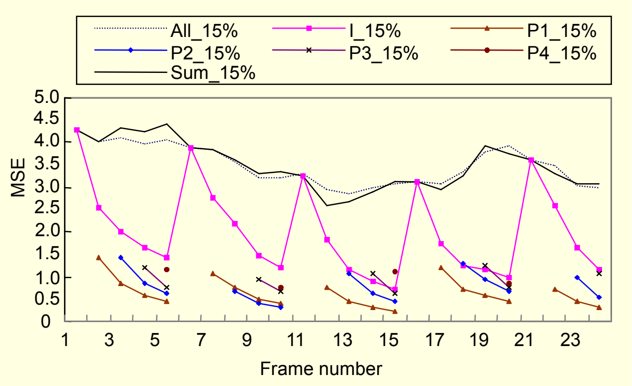
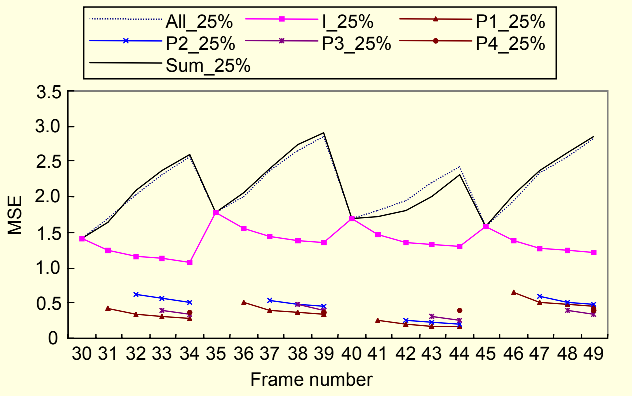


Fig. 2. Each component of the CD distortion for Foreman sequence.

Second, CD with a uniform ratio of bit rate reduction is applied to the Foreman sequence on a frame by frame basis. That is, after FD of all B-frames, the CD operation is applied to each frame individually and the error propagation behavior is investigated as shown in Fig. 3(a). For instance, "I_15%" indicates the result of a 15% rate reduction of the I-frame only, while keeping other coded frames not transcoded. Similarly, "Px_15%" is the result of a 15% rate reduction of the x-th P-frame only, while keeping other coded frames (including all I-frames) not transcoded. In Fig. 3(a), we see that each CD distortion is propagated to subsequent frames in a monotonically decreasing manner (see the plots of I_15%, P1_15%, and so on). Note that the sum of the propagated errors from the past frames and the current error ("sum_15%") approximates the result of CD applied to all frames together



(a) Foreman sequence



(b) Akiyo sequence

Fig. 3. Error propagation behavior of the CD distortion.

("all_15%"). In Fig. 3(b), a 25% rate reduction is applied to the "Akiyo" sequence. Figure 3(b) also shows that the overall distortions can be approximately expressed by the independent sum of the current error and the propagated errors from the previous frames. Therefore, the estimation of CD distortion becomes feasible due to the independent modeling of the error propagation of each frame. Accordingly, as stated in property 2, when CD is applied to multiple frames simultaneously, each CD distortion can be approximated by the sum of the current error and the propagated errors from all the previous frames.

On the other hand, in Figs. 3(a) and (b), the decaying slope of the Akiyo sequence is not steep and, in particular, the CD error of I-frames is critical compared to that of the Foreman sequence. This is mainly due to the characteristics of the coded video sequences, as shown in Table 1. When there is low spatial detail and a small amount of motion in the video as in the class A sequence, the slope decays slowly because each frame is highly correlated in temporal direction and the power of the motion-compensated component is dominant. On the contrary, when there is high spatial detail and a medium amount of motion in the video as in the class C sequence, the slope decays steeply because of some intra-coded blocks. From other experiments, it is observed that when there is medium spatial detail and a medium amount of motion in the video as in the Class B sequence, the slope of the decay is between the other two cases. Therefore, to find a feasible model to describe the error propagation behavior of each frame, it is necessary to effectively exploit the coded block information per frame in the compressed domain.

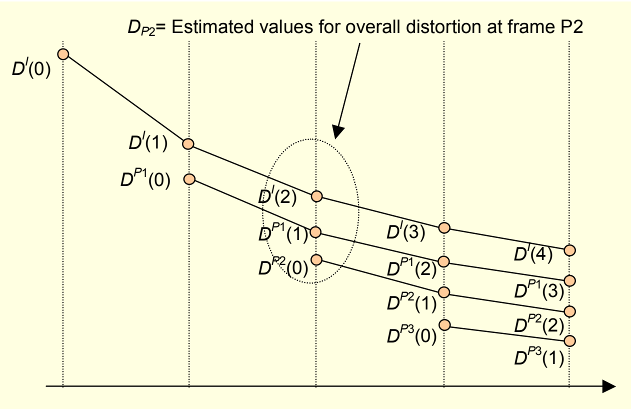


Fig. 4. Adaptive estimation for each propagated distortion within one GOP.

3. An Adaptive Estimation Model for the Overall CD Distortions

From the above simulations, it is observed that the CD distortion is propagated to subsequent frames in a monotonically decreasing manner. That is, the propagation behavior of the CD distortion can be stated as the following property.

Property 3. Let x_j be the decoded value of the m-th pixel in scan order within block $y_{j,i}$, and let x_{j-1} be the motion-compensated component of x_j, that is, the m-th value of block $C_{j,i}(y_{j-1})$ in the same scan order. If the relationship between these two values can be described as a first-order autoregressive [AR(1)] sequence [16] and random MSE distortion $\delta_j(0)$ is added to x_{j-1}, the distortion is propagated to x_j with $\delta_j(1) \le \delta_j(0)$. Extending this relationship to frame (j+2), frame (j+3) and so on, the distortion is propagated to subsequent samples of the decoded sequence in a monotonically decreasing manner: $\delta_j(p) = \zeta\delta_j(p-1)$ for $\zeta \le 1, p \ge 1$.

Proof. An AR(1) process x_j is defined as $x_j = \lambda x_{j-1} + \varepsilon_j$, $0 \le \lambda \le 1, j \ge 0$. Let $\hat{x}_{j-1}$ denote the degraded signal incurred by adding random distortion to x_{j-1}. For a given MSE distortion,

$$\delta_j(0) = E[(x_{j-1} - \hat{x}_{j-1})^2],$$

$$\begin{aligned}\delta_j(1) &= E[(x_j - \hat{x}_j)^2] \\ &= \lambda^2 E[(x_{j-1} - \hat{x}_{j-1})^2] = \lambda^2 \delta_j(0)\end{aligned}$$

are obtained. According to this relationship, $\delta_j(0) \ge \delta_j(1)$ is found. The extension of this proof gives $\delta_j(0) \ge \delta_j(1) \ge \cdots \ge \delta_j(p-1)$ for $p \ge 1$. □

Based on properties 1 and 2, it is shown that the CD distortion can be approximated by the independent sum of the current errors and the propagated errors. Figure 4 shows an adaptive estimation model for each propagated error. In Fig. 4, for example, coded frame indices within one GOP are denoted as the I-frame, the first P-frame, and so on. This pattern is repeated GOP by GOP. By generalizing the expression, an overall distortion of the *j*-th frame is simply described as

$$D_j = D^j(0) + \sum_{p \ge 1} D^{j-p}(p), \tag{10}$$

where the first term is the distortion caused by CD applied to the *j*-th frame and the second term is the propagated distortions from the CD operations of previous frames, as denoted in Fig. 4.

The error propagation behavior of each frame is directly related to the decaying slope of the frame. According to property 3, the second term of (10) is found adaptively frame by frame. For this purpose, let ρ_{j-p+q} denote the decaying slope of the *(j-p+q)*th frame, representing how much the previous distortion $D^{j-p}(q-1)$ contributes to the current one $D^{j-p}(q)$ in a monotonically decreasing manner. By using this parameter, a model describing the propagation behavior of each CD distortion in (10) is proposed in an adaptive manner as

$$D^{j-p}(q) = \rho_{j-p+q} D^{j-p}(q-1), \tag{11}$$

where $D^{j-p}(0)$ means the distortion caused by CD applied to the (*j-p*) th frame. In (11), as *q* increases, $D^{j-p}(q)$ is monotonically decreased as shown in Fig. 3.

As mentioned in the previous experimental results, the decaying slope, ρ_{j-p+q}, is directly dependent on the coded characteristics of frame (*j-p+q*). To find the model parameter, the spatial and temporal activities are measured, based on the statistical characteristics of coded blocks per frame. To take account of these spatial and temporal activities, ρ_{j-p+q} is chosen to be a compound form as follows:

$$\rho_{j-p+q} = \alpha_{j-p+q}\left\{1 + \beta_{j-p+q}\right\} \cdot \gamma_{j-p+q} \quad , \quad \left|\rho_{j-p+q}\right| \le 1. \tag{12}$$

Here, α_{j-p+q} reflects the spatial activity of the (*j-p+q*)th frame, α_{j-p+q} represents the power portion compensated from the (*j-p+q*-1)th frame in the overall power of all inter-coded blocks (denoted the block set as *S*) in the (*j-p+q*)th frame and is measured as

$$\alpha_{j-p+q} = \sum_{i \in S} P^{\text{inter}}_{j-p+q,i} \Big/ \sum_{i \in S} P_{j-p+q,i}\, , \tag{13}$$

where $P^{\text{inter}}_{j-p+q,i}$ and $P_{j-p+q,i}$, which denote the propagated power and the total power of the inter-coded block *i*, respectively, can be simply estimated in the DCT domain. However, in this paper, for simple implementation, the propagated power is obtained by calculating the power portion of the MC block over the total block power at the same block position without using moving information.

Furthermore, to exploit the temporal activity between the (*j-p+q*)th frame and the (*j-p+q*-1)th frame, β_{j-p+q} and γ_{j-p+q} are introduced to reflect the portion of the uncoded block and the portion of inter-coded blocks at the (*j-p+q*)th frame, respectively. Based on several experimental investigations, and to take into account different decay slopes of the propagation error, $\beta_{j-p+q} = (M^{not}{}_{j-p+q}/M)^{0.5}$ and $\gamma_{j-p+q} = (M^{not}_{j-p+q} + M^{mc}_{j-p+q})/M$ are obtained, where M^{not}_{j-p+q} is the number of uncoded blocks in inter-coded macroblocks of frame (*j-p+q*), and M^{mc}_{j-p+q} is the number of MC-coded blocks in frame (*j-p+q*).

4. Simulation Results for the Adaptive Estimation Model for the Overall CD Distortions

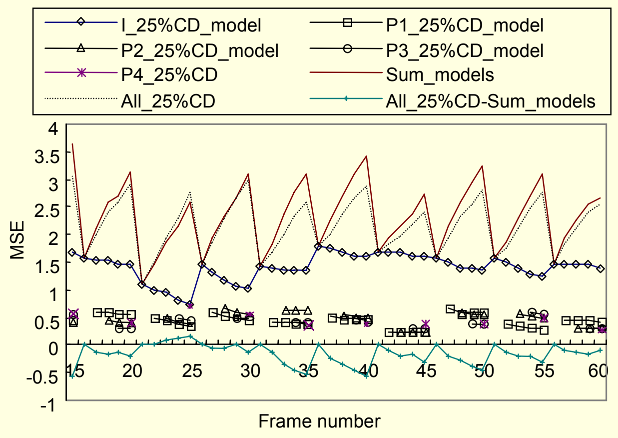
The effectiveness of CD distortion estimation using the proposed model is tested for two quite different sequences, namely, the Akiyo and Stefan sequences. The sequences in CIF format are coded at 1.2 Mbps and 1.5 Mbps with GOP (size = 15, sub-GOP size = 3), respectively. Then, the uniform ratio of bits per frame is truncated. Figure 5 shows the experimental results. In Fig. 5(a), for instance, "I_25%CD_model" indicates the trajectory of the estimated distortion caused by the 25% CD of the I-frame, based on (11), and "Sum_models" denotes the sum of all the individual models as defined in (10). "All_25%CD" is the plot of the actually measured distortion at the 25% CD of all coded frames.

These experimental results show that the distortion estimation model accurately approximates the measured distortion. In particular, when the activity of moving objects is very low as in the Class A sequence, the decay slope of CD errors is not steep. This result is mainly because the Akiyo sequence has a larger percentage of uncoded blocks in inter-coded macroblocks and so the CD errors of I-frames are critical, as shown in Table 1 (Also, the experiment with the Mother and Daughter sequence is very similar to the Akiyo sequence). On the other hand, compared to the Akiyo sequence, since the Stefan sequence has a larger percentage of intra-coded blocks and so the intra-refresh effect is working, the slope of CD errors steeply decays.

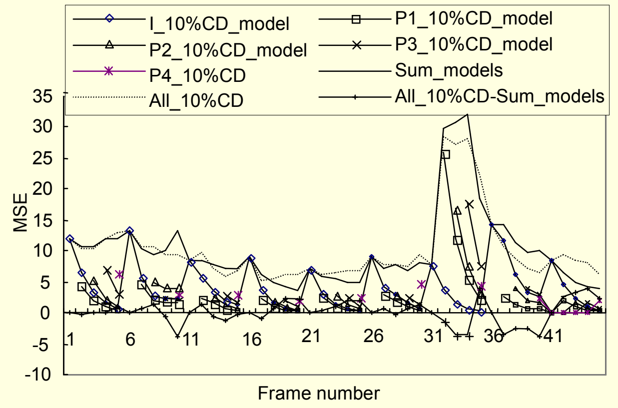
IV. Distortion Control Using Estimation Model

1. Uniform Distortion Allocation Algorithm

As discussed in section III, the loss of high-frequency



(a) Akiyo sequence



(b) Stefan sequence

Fig. 5. Results of the estimation of the CD distortion.

information is subject to non-uniform visual qualities in the reconstructed frames. As time goes on, these errors progressively accumulate, resulting in the reconstructed frames becoming severely degraded. In some applications of the FD-CD transcoder, it is necessary to reduce or exclude the subjective quality fluctuations due to these errors. In this subsection, considering these application areas, let us suppose that there are no explicit rate constraints and CD distortion per frame should be maintained as uniformly as possible.

For this aim, by introducing (1), a uniform distortion allocation (UDA) algorithm is designed on a frame-by-frame basis. First, we assume that an allowable distortion per frame, D_{keep}, is to be allocated for all coded frames. Then, for the j-th frame, the problem is to find a new breakpoint set $\vec{k}=(k_1,k_2,\cdots,k_N)$. The problem formulation is given as

$$\underset{\vec{k}=(k_1,k_2,\ldots,k_M)}{\text{minimize}}\{\sum_{i=1}^{M}(\sum_{k=k_i+1}^{N-1} z_{i,k}{}^{2}+\lambda\sum_{k=0}^{k_i} b_{coeff}^{i,k})\}$$

$$\text{such that} \quad \left|D_{keep}-D_j\right| \le \varepsilon_d , \qquad (14)$$

where $z_{i,k}$ is the dequantized-DCT coefficient for the k-th symbol in the i-th block, and ε_d is the allowable distortion bound. To allocate D_{keep} over all coded frames evenly, an overall distortion propagated from the previous frames should first be estimated. That is, the distortion of the j-th frame, D_j, is estimated by using (7)-(10). Then, the practical CD error of the j-th frame should be controlled to satisfy D_{keep}-D_j. Under the distortion constraint, D_{keep}-D_j, the dropped DCT coefficients within the j-th frame are determined by the LOCD scheme. For an example with the second P-frame within GOP, overall propagation errors are incurred by the CD errors of the I-frames and the first P-frame. Denoting j-2 and j-1 as the index of I-frame and the first P-frame, respectively, the practical CD error to be allocated into the second P-frame is D_{keep}-$(D^{j-1}(1)+D^{j-2}(2))$.

2. Simulation Results for UDA Algorithm

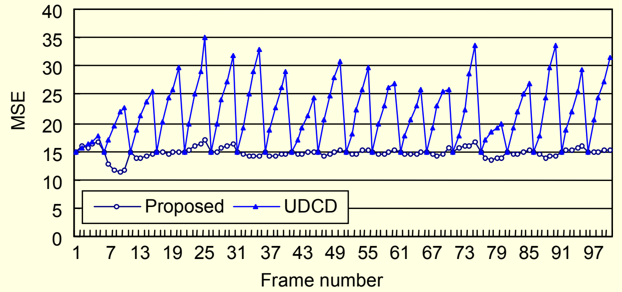
Figure 6 shows the simulation results for the Akiyo and Coast Guard sequences, which were controlled for MSE=15 and MSE=25, respectively. In order to compare the performance of the UDA algorithm, a uniform distortion-based CD (UDCD) scheme was chosen [1], [5], [6], [10], [12], [16], [17]. In the UDCD scheme, without considering the propagated errors of each CD distortion, DCT coefficients corresponding to a constant distortion per frame are dropped. As shown in Fig. 6, by using the proposed UDA algorithm, nearly constant distortions can be controlled over coded frames in the compressed domain. However, in the UDCD scheme, as time goes on, the propagated errors progressively increase and accumulate, resulting in the subsequent frames becoming severely degraded. The experimental results demonstrate that the proposed UDA algorithm is effective for allocating nearly constant distortions into coded frames in DCT domain.

On the other hand, as shown in Fig. 6, the Coast Guard sequence shows a relatively weak control performance compared to that of the Akiyo sequence, particularly, in the region of frames having uncovered objects (namely, frames 21 to 45). This is mainly because moving information is not considered in (13). In every test video sequence, the UDA algorithm is very effective for controlling uniform distortion over coded frames in the compressed domain.

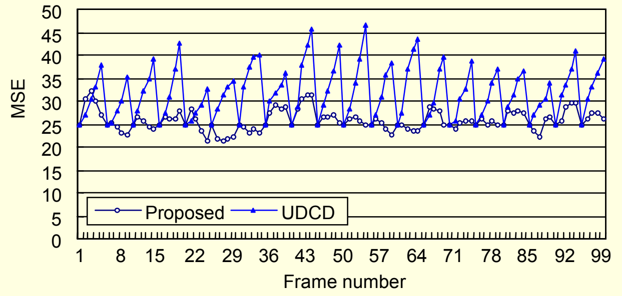
3. Proposed GOP-Based Rate-Distortion Control Algorithm

The challenge of handling (1) in the FD-CD transcoder is finding a way to allocate a given bit budget into the retained frames. Some studies on rate-distortion control in the FD-CD transcoder have focused on the URCD [1], [5], [10], [12]. The operation of URCD is simple.

Based on the target bit rate, a uniform ratio of bits is dropped



(a) Akiyo sequence (MSE=15)



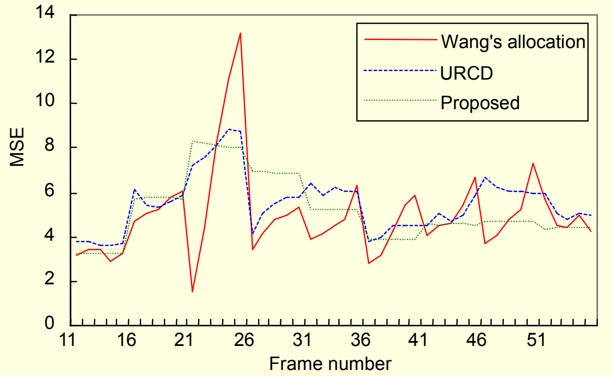
(b) Coast Guard sequence (MSE=25)

Fig. 6. Simulation result of UDA algorithm.

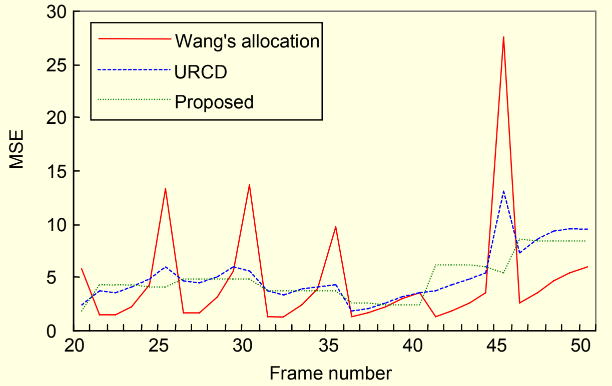
from each frame. Then, the optimal selection of coefficients to be dropped within a frame is achieved by uniformly distributing them into each DCT block, based on (1). However, the URCD scheme suffers from significant visual quality degradations in reconstructed frames. In [11] and [12], Wang used a new global CD rate-allocation scheme to obtain better visual quality. The scheme allocates a larger number of coefficients to be dropped to several frames having strong decoding dependency in the temporal direction. However, this scheme is not adaptive to the characteristics of a given coded-video stream and may result in high fluctuations of decoded visual quality.

In this subsection, the distortion estimation model is exploited to develop a content-adaptive rate-distortion control algorithm in the compressed domain. The main goal of the rate-control scheme presented in this paper is to allocate a given bit-budget into coded frames within one GOP, while keeping CD distortions as constant as possible. For this aim, let s represent an index of the ordered picture coding type within a single GOP, that is, $s \in \{I, P1, P2, \cdots\}$. Then, let us suppose a coded frame with the picture coding type s undergoes transcoding from $r(s)$ into $r_i'(s)$ [$r(s) > r_i'(s)$ with iteration index i]. By using these notations, (1) is modified to develop the GOP-based rate-distortion control as follows.

Step 1. Let us denote R and R' ($<R$) as the number of original bit counts over one GOP (after only FD is applied) and the number of target bit counts over one GOP (after both FD-CD are applied), respectively. As presented in [1], [5], [11], auniform ratio of bits, denoted as $\eta = R'/R = r'(s)/r(s)$, is calculated for each frame. Then, based on (1), a coded frame with the picture coding type s undergoes the CD operation under the rate constraint $r_0'(s) = \eta \cdot r(s)$, where index 0 means initial allocation.



(a) Mother and Daughter



(b) Table Tennis

Fig. 7. Simulation results for the rate control schemes.

Step 2. An overall distortion of the coded frame s, D_s, is estimated by using (10). Based on the estimated distortions, D_s for $s \in \{I, P1, P2, \cdots\}$, the averaged distortion, $D_{avg} = S^{-1}\sum_s D_s$, is found, where S is the number of the frames kept within one GOP. If $|D_{avg} - D_s| \le \varepsilon_d$ for $s \in \{I, P1, P2, \cdots\}$ is satisfied, this routine is stopped; otherwise, go to step 3.

Step 3. Under the averaged distortion, D_{avg}, found in step 2, a newly allocated bit count with the i-th iteration, $r_i'(s)$ for $s \in \{I, P1, P2, \cdots\}$, is re-allocated such that $|D_{avg} - D_s| \le \varepsilon_d$ can be satisfied, based on (14). That is, if $D_s < D_{avg}$, AC DCT coefficients corresponding to D_{avg}-D_s are additionally dropped at frame s. Then, $r_i'(s)$ is reduced by the number of bits dropped. On the contrary, if $D_s > D_{avg}$, $r_i'(s)$ is increased by gathering the coefficients bit counts corresponding to D_s - D_{avg}. This is performed by finely re-allocating D_{avg} into frame s in (14).

Step 4. For the newly re-allocated rates, $r_i'(s)$, $s \in \{I, P1, P2, \cdots\}$ found in step 3, if the condition $\left|\sum r_i'(s) - R'\right| \le \varepsilon_r$ is satisfied, this routine is stopped; otherwise, go to step 5.

Step 5. A newly allocated bit counts with the $(i+1)$th iteration is re-allocated as

$$r_{i+1}'(s) = r_i'(s) + \frac{R' - \sum_s r_i'(s)}{\sum_s r(s)} r(s)$$

and undergoes the CD operation under the rate constraint $r_{i+1}'(s)$. Then, go to step 2.

4. Simulation Results for the Rate-Distortion Control

As in the previous experiments, all B-frames are dropped. Then, the allowable bit-budget is allocated to the retained frames by three algorithms, namely, URCD [10], Wang's allocation [12], and the proposed algorithm. First, the Mother and Daughter sequence coded at 800 kbps is transcoded to 360 kbps. Figure 7(a) shows the simulation results. Since most of the P-frames in the Mother and Daughter sequence consist of inter-coded blocks and low spatial details, the propagation errors cause the severe blurring of successively predicted frames. Wang's allocation distributes more bits into the I-frame, the P1-frame,···, in a decreasing manner. However, the decoded quality fluctuation is very severe, while the averaged distortion is lowest at some frames. Also, URCD suffers from severe smoothing, and CD distortions are also accumulated.

As shown in Fig. 7(b), the Table Tennis sequence coded at 3.0 Mbps is reduced to 1.5 Mbps. Compared with the Mother and Daughter sequence, the Table Tennis sequence has a smaller number of inter-coded blocks [18], so the predictive frames contribute to a smaller propagation error. Accordingly, Wang's allocation is not effective for this kind of coded sequence. In contrast, since URCD treats each coded frame as an intra-coded one, to drop a uniform ratio of bits per frame has an effect on the allocation of nearly uniform distortion per frame. However, URCD fails to adaptively control the propagation errors. From the above simulations, it is evident that the proposed scheme is very effective for controlling rate-distortion over coded frames in the compressed domain, even if the algorithm is designed for the duration of one GOP window.

V. Conclusion

In this paper, to effectively control the distortions incurred by DCT coefficient dropping in the compressed domain, a novel estimation model describing propagation of the CD distortion is proposed. First, essential statistical properties of the CD distortion are investigated along with empirical observations. Several experiments with different compressed sequences have demonstrated the effectiveness of the estimation model. The proposed model enables efficient estimation of the distortion depending on the number of coefficients dropped in the compressed domain. Furthermore, without introducing a full decoding and re-encoding system in the pixel domain, it was demonstrated that the proposed model can easily be extended to allocate nearly uniform distortions among frames and to control an effective rate-distortion control in the compressed domain.

In future works, it will be necessary to analyze the error characteristics by dropping coefficients on a variety of compressed frames or video sequences (for example, B-frames, with scene changes, and so on). Furthermore, we are trying to apply the proposed distortion model to recent codecs (for example, H.264), where entropy coding of DCT coefficients does not follow the traditional zigzag scanning order.

References

[1] J.G. Kim, Y. Wang, and S.F. Chang, "Content-Adaptive Utility-Based Video Adaptation," *Proc. ICME-2003*, Baltimore, Maryland, July 2003.

[2] A. Vetro, C. Christopoulos, and H. Sun, "Video Transcoding Architectures and Techniques: An Overview," *IEEE Signal Processing Mag*., vol. 20, Mar. 2003, pp. 18-29.

[3] H. Sun, W. Kwok, and J.W. Zdepski, "Architectures for MPEG Compressed Bitstream Scaling," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, April 1996, pp. 191-199.

[4] J. Youn, J. Xin, and M.T. Sun, "Fast Video Transcoding Architectures for Networked Multimedia," *IEEE Int. Symp. Circuits and Systems*, Geneva, Switzerland, vol. 4, May 2000, pp. 25-28.

[5] J.G. Kim, J.S. Kim, and H.M. Kim, "Simple Method for Estimating Distortion Caused by DCT Coefficient Drop in Compressed Domain," *IEE Electronics Letters*, Dec. 2004, vol. 40, no. 25, pp. 1581-1582.

[6] J.S. Kim and J.G. Kim, "Efficient Control for the Distortion Incurred by Dropping DCT Coefficients in Compressed Domain," *Proc. PCM-2005*, Cheju, Korea, Nov. 2005.

[7] T.P. Chen and T. Chen, "Rate Shaping for Video with Frame Dependency," *Proc. ICME-2003*, Baltimore, MD, July 2003, pp. I-201 – I-204.

[8] T.P. Chen and T. Chen, "Error Concealment Aware Rate Shaping for Wireless Video Transport," *EURASIP Sig. Proc.: Image Communications, Special Issue on Recent Advances in Wireless Video*, vol. 28, Nov. 2003, pp. 889-905.

[9] A. Eleftheriadis and P. Batra, "Dynamic Rate Shaping of Compressed Digital Video," *IEEE Trans. on Multimedia*, vol. 8,

no. 2, Apr. 2006. pp. 297 -314.

[10] A. Eleftheriadis, “Dynamic Rate Shaping of Compressed Digital Video,” Ph.D. Dissertation, Dept. Elec. Eng., Columbia University, June 1995.

[11] J. Wang and S. Yu, “Dynamic Rate Scaling of Coded Digital Video for IVOD Applications,” *IEEE Trans. on Consumer Electronics*, vol. 44, Aug. 1998, pp. 743-749.

[12] Y. Wang, J.G. Kim, and S.F. Chang, “Content-Based Utility Function Prediction for Real-Time MPEG-4 Video Transcoding,” *Proc. ICIP-2003,* vol.1, Sept. 2003, pp. 189-192.

[13] Y. Wang, J.G. Kim, and S.F. Chang, “MPEG-4 Real Time FD-CD Transcoding,” *Columbia Univ. DVMM Group*, New York, 2003. Tech., Rep. 208-2005-2.

[14] K.R. Rao and P. Yip, *Discrete Cosine Transform; Algorithms, Advantages, Applications – Chapter 2 Definitions and General Properties,* Academic Press, San Diego, 1990.

[15] J. Nieweglowski and P. Haavisto, “A Method for Motion Compensated Video Prediction,” *International Workshop on HDTV*, Ottawa, Canada, Oct. 26-28, 1993.

[16] P. Batra and A. Eleftheriadis, “Analysis of Optimal Dynamic Rate Shaping of Markov-1 Sources,” *Proc. ICIP-2003,* Barcelona, Spain, Sept. 2003.

[17] Y. Liang and Y.P. Tan, “A New Content-Based Hybrid Video Transcoding Method,” *Proc. ICIP-2001*, Oct. 2001, pp. 429-432.

[18] H. Lee, Y. Kim, A.H. Rowberg, and E.A. Riskin, “Statistical Distributions of DCT Coefficients and Their Application to an Interframe Compression Algorithm for 3-D Medical Images,” *IEEE Trans. on Medical Image,* vol. 12, no.3, Sept. 1993, pp. 478-485.

Jin-soo Kim received the BS degree in electronics engineering from Kyungpook National University, Daegu, Korea, in 1991, and the MS and PhD degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Korea, in 1993 and 1998, respectively. From 1998 to 2000, he was with the Business Division of System LSI in Samsung Electronics, where he was involved in the development of MCU chipsets. Since March 2000, he has been a faculty member in the School of Information Communication and Computer Engineering, Hanbat National University, Korea, where he is an associate professor. His research interests include scalable video coding, networked video-rate shaping and adaptation, media convergence, and multimedia re-multiplexing.

Jae-gon Kim received the BS degree in electronics engineering from Kyungpook National University, Daegu, Korea, in 1990, and the MS and PhD degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Korea, in 1992 and 2005, respectively. From 1992 to 2007, he was with Electronics and Telecommunications Research Institute (ETRI), where he was involved in the development of digital broadcasting media services, MPEG-4/7/21 standards and related applications, and convergence media technologies. From 2001 to 2002, he was a staff associate at the Department of Electrical Engineering, Columbia University, New York. He is currently an assistant professor in School of Electronics, Telecommunications and Computer Engineering, Korea Aerospace University, Korea. His research interests include scalable video coding, video adaptation, media convergence, and multimedia applications.

Kwang-deok Seo received the BS, MS, and PhD degrees in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1996, 1998, and 2002, respectively. From Aug. 2002 to Feb. 2005, he was with LG Electronics. Since March 2005, he has been a faculty member of the Computer and Telecommunications Engineering Division, Yonsei University, Gangwon, Korea, where he is an assistant professor. He has over 30 pending or issued patents and has published over forty papers in the areas of multimedia coding, multimedia signal processing, and multimedia communication systems. His current major research interests include scalable video coding, protocol design for scalable video transport, and scalable video streaming system design. He is a member of KICS, IEEE, and IEICE.

Mong-han Yun received the BS and MS degrees in electronics engineering from Korea Aerospace University, Korea, in 1987 and 1990, respectively. From 1990 to 2003, he was with the ADD (Agency for Defense Development) in Korea. Since 2005, he has been working towards a PhD degree in the Department of Multimedia Engineering at Hanbat National University. His research interests include mobile communications, visual communications, and image signal processing.