

Efficient Control for the Distortion Incurred by Dropping DCT Coefficients in Compressed Domain

Jin-Soo Kim^{*}, Mong-Han Yun^{**}, Jong-Kab Park^{***}

ABSTRACT

The primary goal of this paper is to facilitate the rate-distortion control in compressed domain, without introducing a full decoding and re-encoding system in pixel domain. For this aim, the error propagation behavior over several frame-sequences due to DCT coefficients-drop 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. Experimental results show that the proposed model allows us to effectively control rate-distortions into coded-frames over different kinds of video sequences.

Keywords: Coefficient Dropping, Video Transcoding, Distortion Estimation

1. INTRODUCTION

Recently, transcoding based on dropping some parts of bitstream in compressed domain is considered mainly due to the low computational complexity and simple implementation issues.[1,2] This paper deals with a FD (frame-dropping) - CD (coefficient-dropping) scheme that provides the trade-offs between spatial and temporal qualities as well as extending the range of rate reduction. By combining frame-dropping and DCT coefficient-dropping, it is possible to simply adapt the bit rate of a pre-coded video to dynamic available bandwidth, especially in streaming applications. However, the FD-CD transcoder is subject to drift due to the loss of high-frequency information.[2-4]

※ Corresponding Author : Jin-Soo Kim, Address : (300-719) san 16-1, Deongmyeong-dong, Yuseong-gu, Daejeon, Korea, TEL : +82-42-821-1147, FAX : +82-42-821-1595 E-mail : jskim67@hanbat.ac.kr

Receipt date : July. 11, 2005, Approval date : Sep. 30, 2005

^{*} Dept. of Multimedia Engineering, Hanbat National University.

^{**} Dept. of Multimedia Engineering, Hanbat National University. (E-mail : monghan@pinetelecom.com)

^{***} Dept. of Multimedia Engineering, Hanbat National University. (E-mail : jkpark@kopo.or.kr)

※ This research was financially supported by Hanbat National University in the year of 2004.

As time goes on, this error progressively increases, resulting in the reconstructed frames becoming severely degraded. Therefore, in the FD-CD scheme, it is necessary to reduce or minimize the visual quality fluctuations due to these drift errors in compressed domain.[5]

Some conventional works in FD-CD transcoder has focused on the optimal selections of the amount of dropped coefficients within a frame. But, drift is also accumulated and often results in significant visual quality degradation. In [2], each coded DCT coefficient is treated as intra-coded one and its propagated/accumulated errors are ignored.[3,4] In this paper, we analyze the drift error characteristics incurred by the CD and we propose an effective estimation model that, adaptively, describes well the characteristics of propagation errors. Furthermore, based on this model, how to allocate distortions into coded frames in compressed domain is developed.

2. THE STATISTICAL PROPERTIES AND ESTIMATION MODEL OF CD ERROR

2.1 The Statistical Properties of CD Errors

In this paper, it is assumed that only a con-

tiguous string of DCT coefficients at the end of each block is dropped.[2,3] Let $y_j = \{y_{j,i}\}_{i=1,\dots,M}$ denote the frame j having M blocks, where $y_{j,i}$ is the decoded block i of frame j . Based on this notation, let \hat{y}_{j-1} and \hat{y}_j represent the decoded frame $j-1$ and j , respectively, degraded by CD operation. Then, decoded blocks with and without CD can be expressed as follows:

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

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 decoded prediction error. Thus, the MSE (Mean Square Error) distortion of block i of frame j , $D_{j,i}(k_i)$, when coefficients of which the scan order is greater than k_i are dropped (that is, k_i is a new breakpoint of the block i), is given by

$$\begin{aligned} D_{j,i}(k_i) &= \frac{1}{N} \|y_{j,i} - \hat{y}_{j,i}\|^2 = \frac{1}{N} \|C_{j,i}(y_{j-1}) - C_{j,i}(\hat{y}_{j-1}) + e_{j,i} - \hat{e}_{j,i}\|^2 \\ &= \frac{1}{N} \left\{ \sum_{k=0}^{N-1} A_{j,i}^2(k) + 2 \sum_{k=k_i}^{N-1} A_{j,i}(k) E_{j,i}(k) + \sum_{k=k_i}^{N-1} E_{j,i}^2(k) \right\} \end{aligned} \quad (2)$$

where N is the block size and $A_{j,i}(k) = DCT\{C_{j,i}(y_{j-1}) - C_{j,i}(\hat{y}_{j-1})\}$, $E_{j,i}(k) = DCT\{e_{j,i}\}$, $\hat{E}_{j,i}(k) = DCT\{\hat{e}_{j,i}\}$, $k = 0, \dots, N-1$. Accordingly, it is noted that the CD distortion involves not only the propagated errors and the current error, but the correlated error as well.

Property 1. The MSE in the j -th frame can be approximated to

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

That is, the correlated error is negligible and thus, the distortion of the j -th frame is determined by the sum of the current error and the propagated errors.

Proof The proof is straightforward based on the fact that $A_{j,i}(k)$ and $E_{j,i}(k)$ are statistically independent and their expected values are zero for $NM \gg 1$ [6].

Property 2. When CD is applied to multiple cod-

ed-frames, together, the overall distortion of each frame can be expressed by the sum of the current error and the propagated errors from the CD operations of previous frames.

Proof The proof is similar to that of Property 1.

2.2 CD Control and Simulation Results for CD Errors

There may be many CD control algorithms.[2-4] In order to keep a uniform quality within each kept frames and to result in more comfortable spatial quality, in this paper, we assume that the defined CD operations have to evenly distribute the dropped coefficients in the spatial range. Accordingly, we consider only Lagrangian Optimization CD (LOCD), which tries to find an optimal truncation point for each block within an optimization window of one frame.[2,3] This is accomplished by using Lagrangian search to minimize the distortion caused by the CD. For the j -th frame with bit budget B_{keep} , the problem is to find a new breakpoints set $\vec{k} = (k_1, k_2, \dots, k_M)$. The problem is formulated as follows:

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^M \sum_{k=k_i+1}^{N-1} z_{i,k}^2 \quad \text{such that} \\ &\sum_{i=1}^M \sum_{k=1}^{k_i} b_{coeff}^{i,k} \leq B_{keep}^i \end{aligned} \quad (3)$$

where $z_{i,k}$ and $b_{coeff}^{i,k}$ is the dequantized-DCT coefficient and the amount of coded bits for the k -th symbol in the i -th block, respectively.

Fig. 1 compares each component of the CD distortion. In this experiment, "Foreman" sequence coded 1.5Mbps in CIF is transcoded to 800kbps by CD with uniform rate reduction among frames and FD of all B-frames dropping[3]. This result makes sure that the correlated term can be ignored while the current CD error and the propagated errors are dominant. So, Property 1 is very useful.

Additionally, Fig. 2 shows the simulation results that CD is applied to each frame individually to in-

investigate the error propagation behavior. In Fig. 2, "X_15%" indicates the result of 15% rate-reduction of the X-frame, only while keeping other coded-frames not transcoded. It is noted that the sum of the propagated errors from the past frames and the current error ("sum_15%") is well approximated to the result of CD applied to all frames together ("all_15%"). Therefore, the estimation of CD distortion becomes feasible due to the modeling of error propagation of each frame, independently. 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 the CD operations of all previous frames.

2.3 Simple Estimation Model for the Overall CD Distortions

From the above experiments, it is observed that the CD distortion is propagated to subsequent frames in a monotonically decreased manner. If the relationship between $y_{j,i}$ and $C_{j,i}(y_{j-1})$ can be described as a first-order autoregressive [AR(1)] sequence, the propagation behavior of the CD distortion can be stated as follows:

Property 3. If the relationship between the m -th pixel value of $y_{j,i}$ and the m -th MC value of $C_{j,i}(y_{j-1})$ in scan order can be described as a first-order autoregressive [AR(1)] sequence and

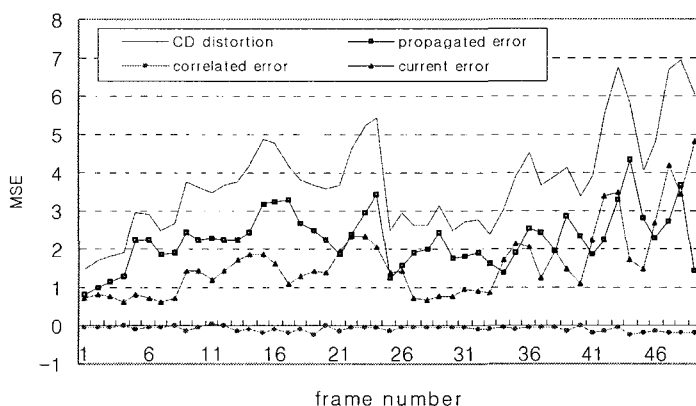


Fig. 1. Each component of the CD distortion for "Foreman".

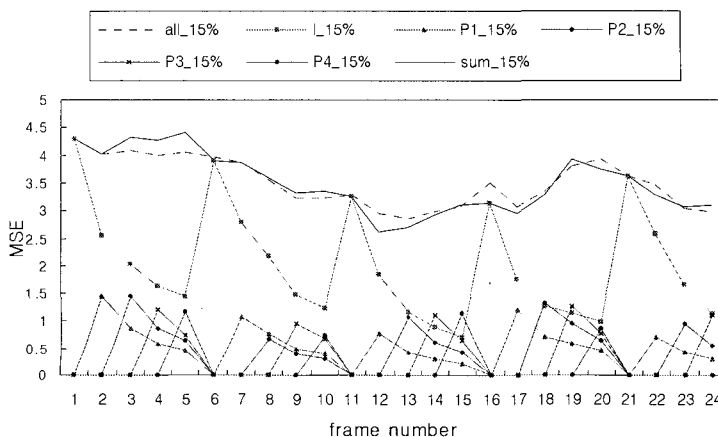


Fig. 2. Error propagation behavior of the CD distortion for "Foreman".

random MSE distortion $\delta(0)$ is added to $C_{j,i}(y_{j-1})$, the distortion is propagated to subsequent samples in a monotonically decreased manner, i.e., $\delta_j(p) = \zeta \delta_j(p-1)$, $\zeta \leq 1$, $p \geq 1$.

Proof) An AR(1) process x_{j+1} is defined as $x_{j+1} = \lambda x_j + \epsilon_{j+1}$, $0 \leq \lambda \leq 1, j \geq 1$. Let \hat{x}_j denote the degraded signal incurred by adding random distortion to x_j [7]. For a given MSE distortion $\delta_j(0) = E[(x_j - \hat{x}_j)^2]$,

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

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

According to Property 1 and 2, an overall distortion of the j -th frame is simply approximated as follows:

$$D_j = D^j(0) + \sum_{p \geq 1} D^j(p) \quad (4)$$

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. From the Property 3, a model describing propagation behavior of each CD distortion is proposed in a recursive and adaptive manner as follows:

$$D^{j-p}(q) = \rho_{j-p+q} D^{j-p}(q-1), p \geq q \geq 1 \quad (5)$$

where $D^{j-p}(0)$ means the distortion caused by CD applied to the $(j-p)$ -th frame. The model parameter ρ_{j-p+q} is dependent on the coded-frame $(j-p+q)$ and is introduced to represent how much the previous distortion $D^{j-p}(q-1)$ contributes to the current one $D^{j-p}(q)$ in a monotonically decreased manner as follows:

$$\rho_{j-p+q} = \alpha_{j-p+q} \{1 + \beta_{j-p+q}\} \gamma_{j-p+q} |\rho_{j-p+q}| \leq 1 \quad (6)$$

Here α_{j-p+q} expresses the power portion compensated from the $(j-p+q-1)$ -th frame in the overall power of all inter-coded blocks $\forall i \in S$ in the $(j-p+q)$ -th frame, β_{j-p+q} and γ_{j-p+q} are introduced

to reflect the portion of the not-coded block and the portion of inter-coded blocks at the $(j-p+q)$ -th frame, respectively. α_{j-p+q} is given

$$\alpha_{j-p+q} = \sum_{i \in S} P_{j-p+q,i}^{inter} / \sum_{i \in S} P_{j-p+q,i} \quad (7)$$

where P_{j-p+q}^{inter} and P_{j-p+q} , which denote the propagated power and the total power of the inter-coded block i , respectively, can be simply estimated in the DCT domain. Based on our empirical investigation, to take into account different decaying slopes of the propagation error, $\beta_{j-p+q} = (M_{j-p+q}^{not}/M)^2$ is obtained, where M_{j-p+q}^{not} is the number of not-coded block in inter-coded macroblocks of frame $(j-p+q)$, and $\gamma_{j-p+q} = (M_{j-p+q}^{not} + M_{j-p+q}^{mc})/M$, where M_{j-p+q}^{mc} is the number of MC-coded blocks in frame $(j-p+q)$.

2.4 Simulation Results for Simple Estimation Model

The effectiveness of the CD distortion estimation using the proposed model is tested for two quite different sequences, i.e. "Akiyo" and "Stefan" sequences. The "Akiyo" and "Stefan" sequences in CIF format are coded at 1.2 Mb/s and 1.5 Mb/s with GOP (size = 15, sub-GOP size = 3), respectively. Then, uniform ratio of bits per frame is truncated. Fig. 3 and Fig. 4 show the experimental results. In Fig. 3, for instance, "X_25%CD_model" indicates the trajectory of the estimated distortion caused by the 25% CD of the X-frame, based on (5), and "Sum_models" denotes the sum of all the individual models as defined in (4). "All_25%CD" is the plot of the practically measured distortion at the 25% CD of all coded frames for "Akiyo". These experimental results show that the estimated distortion model well approximates the measured distortion. Particularly, it is noted that, in the case that the activity of moving objects is very low like "Akiyo" sequence, the decaying slope of CD errors is not steep. This result is mainly due to the fact that "Akiyo" sequence has larger percentage of

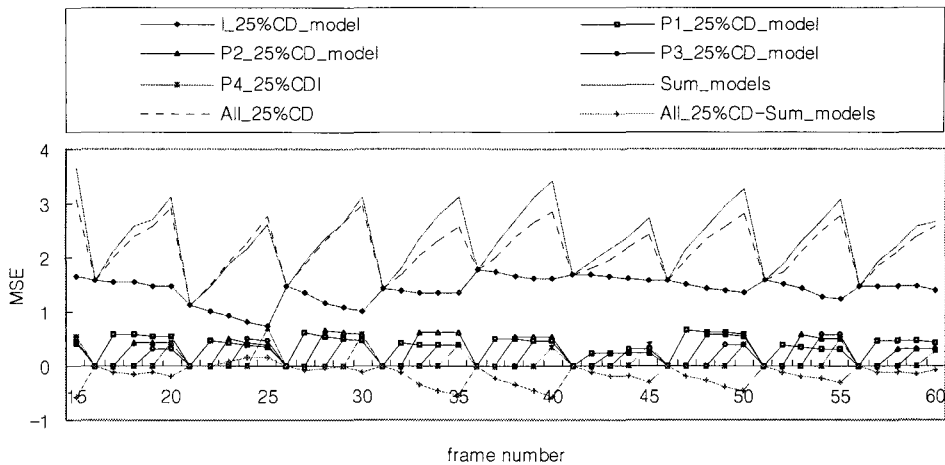


Fig. 3. The results of the estimation of the CD distortion for "Akiyo".

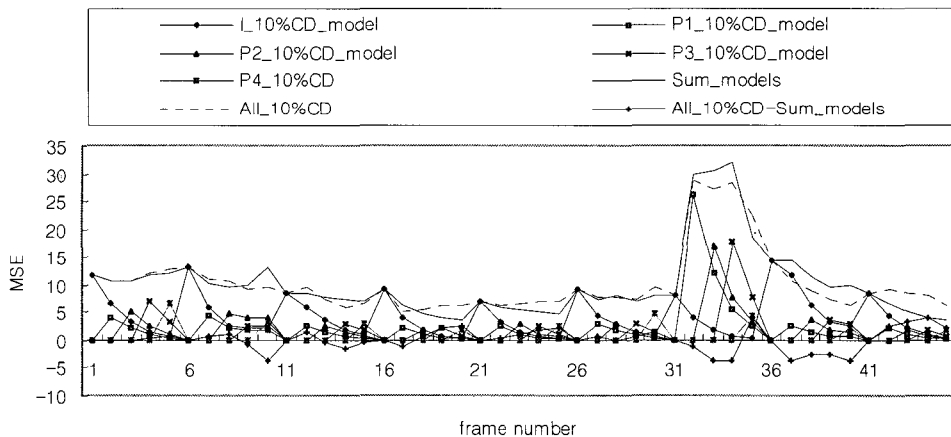


Fig. 4. The results of the estimation of the CD distortion for "Stefan".

not-coded blocks in inter-coded macroblocks and so the CD errors of I-frame are critical. On the other hand, since "Stefan" sequence has larger percentage of intra-coded blocks and so intra-refresh effects are working, the slope of CD errors is steeply decaying.

3. RATE-DISTORTION CONTROL USING PROPOSED ESTIMATION MODEL

3.1 A GOP-based Rate-Distortion Control Algorithm for a Pre-coded Video

The challenge of handling (3) in FD-CD trans-

coder is how to allocate a given bit budget into the kept frames in compressed domain. Some conventional works on the rate-distortion control in FD-CD transcoder has focused on the uniform rate-based CD (URCD).[2-4] URCD operates simply as follows: based on the target bit rate, a uniform ratio of bits is dropped from each frame. Then, the optimal selections of dropped coefficients within a frame are achieved by uniformly distributing into each DCT block, based on (3). In [3], Wang has used a new CD rate-allocation scheme to obtain better visual quality. The scheme allocates the larger number of coefficients to be dropped into several frames having strong decoding

dependency in the temporal direction. In this paper, the distortion estimation model is applied to develop a content-adaptive rate-distortion control algorithm. That is, the main goal of rate-control scheme is to allocate a given bit-budget into coded-frames within one GOP, while keeping overall distortions per frame to be as constant as possible. For this aim, let's represent an index of the ordered picture coding type within single GOP, i.e. $s \in \{I, P1, P2, \dots\}$. Then, let us suppose a coded frame with the picture coding type s undergoes transcoding from $r(s)$ into $r'(s)$ [$r(s) > r'(s)$ with iteration index i]. By using these notations, (3) is modified to develop the GOP-based rate-distortion control as follows:

Step 1. (Uniform ratio of bit reduction) Let us denote R and $R' (< R)$ as the amount of original bit counts over one GOP (after only FD is applied) and the amount of target bit counts over one GOP (after both FD-CD are applied), respectively. As presented in [3], a uniform ratio of bits, denoted as $\eta = \frac{R'}{R} = \frac{r'(s)}{r(s)}$, is calculated for one GOP window. Then, based on the (3), 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 bit allocation.

Step 2. (Distortion estimation) Under the previous bit allocation, an overall distortion of the coded frame s , D_s , is estimated by using (4). Based on the estimated distortions, D_s for $s \in \{I, P1, P2, \dots\}$, the averaged distortion, $D_{avg} = \frac{1}{S} \sum_s D_s$, is found where S is the number of the kept frames within one GOP. If $|D_{avg} - D_s| \leq \epsilon_d$ for $s \in \{I, P1, P2, \dots\}$ is satisfied, this routine is stopped. Otherwise, go to Step 3.

Step 3. (Distortion allocation) By using the averaged distortion, D_{avg} , found in Step 2, a newly allocated bit count with the i -th iteration, $r'_i(s)$ for $s \in \{I, P1, P2, \dots\}$, is re-allocated such that

$|D_{avg} - D_s| \leq \epsilon_d$ can be satisfied, based on (3).

Step 4. (Rate check) For the newly re-allocated rates, $r'_i(s)$, $s \in \{I, P1, P2, \dots\}$ found in Step 3, if the condition $\left| \sum_s r'_i(s) - R \right| \leq \epsilon_r$ is satisfied, this routine is stopped. Otherwise, go to Step 5.

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

$$r'_{i+1}(s) = r'_i(s) + \frac{R - \sum_s r'_i(s)}{\sum_s r(s)} r(s) \text{ and undergoes}$$

the CD operation under the rate constraint $r'_{i+1}(s)$. And then, go to Step 2.

3.2 Simulation Results for the Proposed Rate-Distortion Control Algorithm

As conducted in the previous experiments, all B-frames are dropped and then, the allowable bit-budget is allocated to the kept frames by three algorithms, i.e. URCD[2], Wang's allocation[3], and the proposed algorithm. First, "Akiyo" sequence coded at 1.2Mb/s is transcoded to 500kb/s. Fig. 5 shows the simulation results where "Before_CD" represents the amount of bits per frame before dropping coefficients. Since all P-frames of "Akiyo" sequence consist of inter-coded blocks, the propagation errors bring the severe blurring of successively predicted frames. "Wang's allocation" distributes more bits into I-frame, P1-frame, ..., in decreasing manner. But, it is noted that decoded quality fluctuations is very severe, while the averaged distortion is lowest at some frames. Also, URCD suffers from severe smoothing and the CD distortions are also accumulated.

In Fig. 6, "Foreman" sequence coded at 1.5Mb/s is reduced to 800kb/s. Compared with "Akiyo" sequence, "Foreman" sequence has smaller number of inter-coded blocks and so the predictive frames do contribute to a smaller drift error. Accordingly, "Wang's allocation" is not effective to this kind of coded sequence. In contrast with this scheme, since

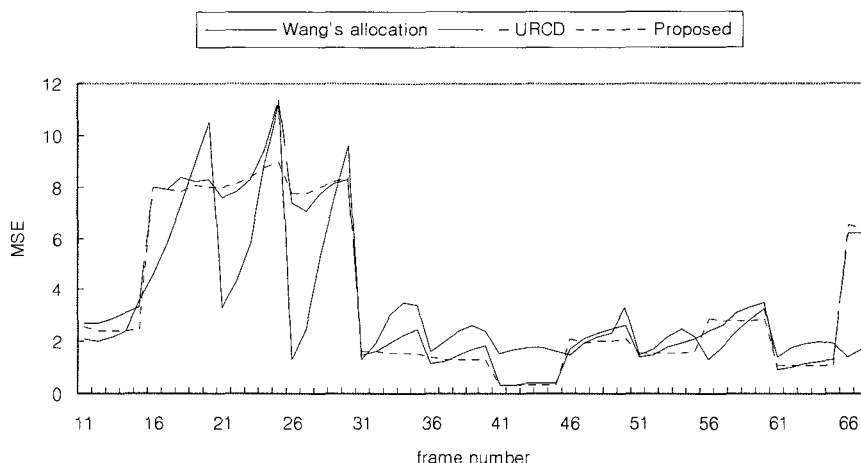


Fig. 5. MSE per frame for "Akiyo".

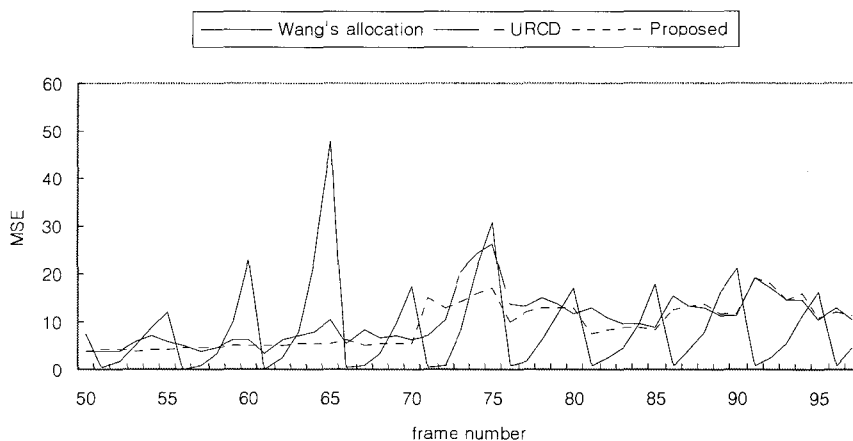


Fig. 6. MSE per frame for "Foreman".

URCD treats each coded frame as intra-coded one, it is found that to drop a uniform ratio of bits per frame has an effect on the allocation of nearly uniform distortion per frame. But, URCD does not fail to adaptively control the drift errors. From the above simulations, it is evident that the proposed scheme is very effective, even if the algorithm is designed for the duration of one GOP window.

4. CONCLUSIONS

In this paper, we have focused on the trans-coding based on dropping some parts of bitstream

in compressed domain. To develop the estimation model, essential statistical properties of the CD distortion are investigated along with empirical observations. Several experiments with different compressed sequences have shown the effectiveness of the estimation model. Without introducing a full decoding and re-encoding system in pixel domain, it is shown that the proposed model can be easily extended to find how to allocate nearly uniform distortions among frames and how to control an effective rate-distortion control in compressed domain. Additionally, it is expected that, by exploiting additional coded-information such as

motion vectors, quantization parameters for each coded frame, much better performance can be achieved.

5. REFERENCES

- [1] A. Vetro, "Video Transcoding Architectures and Techniques: An Overview," *IEEE Signal Processing Mag.* Vol. 20, pp. 18-29, 2003.
- [2] A. Eleftheriadis, "Dynamic Rate Shaping of Compressed Digital Video," *Ph.D. dissertation, Dept. Elec. Eng., Columbia University*, 1995.
- [3] Y Wang, J.G. Kim, and S.F. Chang, "Content-based Utility Function Prediction for Real-time MPEG-4 Video Transcoding," *ICIP'03*, Vol. 1, pp. 189-192, 2003.
- [4] J.G. Kim, Y. Wang, and S.F. Chang, "Content-Adaptive Utility-based Video Adaptation," *ICME-2003*, Baltimore Maryland, 2003.
- [5] J.S.Kim and et al, "A Distortion Control Scheme for Allocating Constant Distortion in FD-CD Video Transcoder," *ICME-2004*, Taipei Taiwan, 2004.
- [6] T. Eude and et al, "On the Distribution of the DCT Coefficients," *Proc. of ICASSP-94*, Vol. 5, pp. 365-368, 1994.
- [7] P. Batra and A. Eleftheriadis, "Analysis of Optimal Dynamic Rate Shaping of Markov-1 Sources," *Proc. Of ICIP-2003*, Barcelona, Spain, 2003.



Jin-Soo Kim

He received the B.S. degree in Electronics Engineering from Kyungpook National University, in 1991, the M.S. and Ph.D. degrees in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), in 1993 and 1998, respectively. From 1998 to 2000, he was a senior researcher at System LSI Sectors of Samsung Electronics. Since 2000, he has been a faculty member of Hanbat National University, where he is now an associate professor in Department of Multimedia Engineering. His research interests are in the areas of image processing, video coding, and visual transmission.



Mong-Han Yun

He received the B.S. and the M.S. degrees in Electronics Engineering from Hankuk Aviation University, in 1987 and 1990, respectively. Since 2005, he is working towards the Ph.D. degree in Multimedia Engineering at Hanbat National University. His interests are the area of visual communications, mobile communications, and signal processing.



Jong-Kab Park

He received the B.S. degree in Computer Engineering from Korea National Open University, in 1995, in a M.S. in Electronics Engineering from Kangwon University in 1997. Since 2005, he is working towards the Ph.D. degree in Multimedia Engineering at Hanbat National University. His interests are the area of visual communications and multimedia authoring.