Hybrid Frame-Recursive Block-Based Distortion Estimation Model for Wireless Video Transmission

Werayut Saesue [#1], Jian Zhang [*2], Chun Tung Chou [#3]

School of Computer Science and Engineering, The University of New South Wales, Sydney 2052, Australia

* *National ICT Australia, Sydney 1466, Australia*

[1] wsae207@cse.unsw.edu.au, [2] jian.zhang@nicta.com.au, [3] ctchou@cse.unsw.edu.au

Abstract—In wireless environments, video quality can be severely degraded due to channel errors. Improving error robustness towards the impact of packet loss in error-prone network is considered as a critical concern in wireless video networking research. Data partitioning (DP) is an efficient error-resilient tool in video codec that is capable of reducing the effect of transmission errors by reorganizing the coded video bitstream into different partitions with different levels of importance. Significant video performance improvement can be achieved if DP is jointly optimized with unequal error protection (UEP). This paper proposes a fast and accurate frame-recursive block-based distortion estimation model for the DP tool in H.264.AVC. The accuracy of our model comes from appropriately approximating the error-concealment cross-correlation term (which is neglected in earlier work in order to reduce computation burden) as a function of the first moment of decoded pixels. Without increasing computation complexity, our proposed distortion model can be applied to both fixed and variable block size intra-prediction and motion compensation. Extensive simulation results are presented to show the accuracy of our estimation algorithm.

I. INTRODUCTION

Transmitting video packets over error-prone wireless channels gives rise to many research challenges. It is widely known that the quality of wireless link varies over time and space due to multi-path fading, shadowing, inter-symbol interference (ISI) and noise. These time-varying and error-prone characteristics of wireless channel reduce the performance of video applications as the errors resulted from packet loss or delay tend to propagate to subsequent frames of the video sequence. Due to constrained bandwidth and limited Quality-of-Service (QoS), high video compression algorithms are preferred in order to diminish traffic congestion in the networks. However, with an increase in video compression efficiency, the quality of decoded video can be severely degraded by a single bit of error as the highly compressed video got rid of both spatial and temporal redundancy from the video data, making it extremely vulnerable to corrupted bits. Although Rate-Distortion (RD) optimization has been widely used to cope with network traffic congestion, the bit error rate (BER) in the wireless channel is still high and can lead to low video quality. Therefore, channel errors must be taken into account while considering the error resilience mechanisms in the video encoder in order to provide acceptable decoded video quality.

Recently, many distortion estimation algorithms have been proposed [1]–[6]. These research works have been directly or indirectly inspired by the Recursive Optimal Per-pixel Estimation (ROPE) algorithm [1]. ROPE accurately estimates the decoded video distortion by recursively calculating the first and second-order moments at pixel level taking into account both quantization distortion and transmission-error distortion. The concept of inter/intra mode switching is also presented in this work. The algorithm is then extended in [2] where sub-pel motion compensation is included. They also proposed the pixel-based correlation approximation approach where mean and standard deviation are available. Clearly, the pixel recursive distortion model increases computation significantly, hence it is not suitable for real-time video applications where the delay is the most critical concern and must be minimized.

In [3], the authors argued that the theoretically derived distortion model is unable to describe the behavior of transmitted video signal accurately due to simplifications of model assumptions. They instead proposed an empirical distortion model by employing curve fitting approach. Their framework consists of both source and channel coding such as forward error correction (FEC) and interleaving. Although the distortion model symbolizes the real-world scenarios, it requires some subsets of measured data points which make it less realistic for video applications that do not have feedback channel.

MB-based distortion model is introduced in [4]. By minimizing average distortion at a given rate, the coding mode i.e. intra/inter/skip and synchronization markers are selected individually for each block. The concealed block distortion is estimated by weighting the overlapped neighbor blocks distortion in the previous frame pointed by motion vector (MV) of the concealed block. The weighting factors are calculated from the length of the MVs. However, the model is limited to the block size of 16x16 and 16-pixel search range only. The work in [5] develops the MB level recursive formula to estimate the video distortion at the decoder based on corresponding block. The distortion of each intra-coded block is calculated from the average of both left and upper neighbors of the current block. Similarly, inter-coded block's distortion is derived from the motion-compensated block in the reference frame given that the referenced-block can be either lost or received at the decoder depends on the packet loss probability.

Wang et al. [6] proposed a model that allows the average frame distortion for a video sequence to be estimated through

MMSP 2008

a recursive formula based on a parameter called error prop-agation factor α. The model is able to estimate the distor-tion of decoded video employing either constrained or non-constrained intra-prediction, sub-pel motion compensation and deblocking filter. In addition, their algorithm allows the lost slices to be concealed using temporal concealment approach. Eventually, the error propagation factor for each case of the encoding configurations can be derived by computing the least square fitting method on sample data obtained via feedback channel.

Nevertheless, Wang's model [6] has several drawbacks that will reduce its usability in the real-world applications. Specifically, [6] reduces the computation complexity of its distortion model by assuming that the error-concealment cross-correlation is negligible. Although this assumption may be ap-propriate for pixel-level and MB-level distortion model, where the computation is exceptionally high, it may not hold for the frame-level distortion model because neglecting the error-concealment correlation term may result in highly inaccurate end-to-end distortion prediction. In order to overcome this aberration in the previously proposed distortion estimation model, we proposed a statistically-derived frame-recursive channel distortion model for data partitioning which incorpo-rates an accurate error-concealment correlation approximation. It is known that data partitioning can increase error robustness in wireless environments [7]. The work presented here is capa-ble of estimating channel-error distortion of transmitted video where constrained intra-prediction and variable block size predictive coding are employed on the encoder i.e. H.264/AVC codec [8]. The rest of this paper is organized as follows: Section II presents our proposed channel distortion estimation model and the corresponding cross-correlation approximation method; Section III verifies the proposed estimation model and discusses the experimental results compared to the model; Section IV summarizes the important results and concludes the paper.

II. CHANNEL DISTORTION ESTIMATION MODEL FOR DATA PARTITIONING

In this section, we first introduce the assumptions and notation of our proposed channel distortion model. We then describe the derivation of our estimation model for both intra-coding and inter-coding mode by associating the MB intra-rate β and cross-correlation approximation.

A. Overview of the Model

Generally, each video sequence is coded into a series of group of pictures (GOP) and each coded GOP in H.264/AVC standard [8] begins with a specially coded picture called IDR (Instantaneous Decoding Refresh) frame. An IDR frame is in fact an intra-coded frame with a synchronisation marker that stops error propagation in the video sequence by prohibiting all the pictures prior to this frame to be referenced by all subsequent pictures in the stream. We assume that the IDR frames must be successfully received without any bit errors in the coded data, this assumption is to ensure that the transmitted

GOP is decodable. We further assume that each video frame is coded into one single independently decodable slice and each slice can has up to three partitions i.e. data partition A (DP-A), B and C. Each partition is packetized into one single RTP packet.

We first define a number of notation for our model. Let D_n denote the average over-all-pixels distortion at frame n. Let f_n^i and $\hat{f}_n^i$ denote the original value and its encoder reconstructed value of pixel i in frame n, respectively. The decoder recon-structed value (after error concealment) is represented by $\tilde{f}_n^i$. Assuming Mean-Squared-Error (MSE) is used, the total frame distortion D_n is thus

$$D_n = E\left\{(f_n^i - \tilde{f}_n^i)^2\right\} \tag{1}$$

$$= E\left\{(f_n^i - \hat{f}_n^i + \hat{f}_n^i - \tilde{f}_n^i)^2\right\}$$

$$= E\left\{(f_n^i - \hat{f}_n^i)^2\right\} + E\left\{(\hat{f}_n^i - \tilde{f}_n^i)^2\right\}$$

$$+ 2E\left\{(f_n^i - \hat{f}_n^i)(\hat{f}_n^i - \tilde{f}_n^i)\right\} \tag{2}$$

$$\approx D_{e,n} + D_{c,n} \tag{3}$$

where $D_{e,n} = E\left\{(f_n^i - \hat{f}_n^i)^2\right\}$ is the encoder distortion caused by quantization and $D_{c,n} = E\left\{(\hat{f}_n^i - \tilde{f}_n^i)^2\right\}$ is chan-nel distortion caused by transmission loss. According to [9], the terms $f_n^i - \hat{f}_n^i$ and $\hat{f}_n^i - \tilde{f}_n^i$ are pretty much uncorrelated so the correlation term in Eq. (2) can be ignored. We know from previous work [1]-[5] that the encoder distortion $D_{e,n}$ can be accurately estimated from the available information on the en-coder side. However, determining the channel distortion $D_{c,n}$ poses many research challenges due to frame dependencies and error characteristics of wireless channel. Our work focuses only on estimating the channel-induced distortion $D_{c,n}$.

In this paper, we adopt both MB-based and frame-recursive distortion model to jointly estimate the average frame distor-tion. Our hybrid algorithm individually calculates the distor-tion of intra-coded MBs (I-MBs) and inter-coded MBs (P-MBs) composed in each frame. The ratio of I-MBs over the total number of coded MBs is called the *intra-rate* of frame n and is referred by β_n. The recursion formula is given by the weighted mean of I-MBs distortion ($^I D_{c,n}$) and P-MBs distortion ($^P D_{c,n}$), as follows:

$$D_{c,n} = {}^I D_{c,n}\,\beta_n + {}^P D_{c,n}\,(1 - \beta_n) \tag{4}$$

We will derive the channel distortion for I-MBs and P-MBs separately in subsequent sections.

B. I-MB Distortion

The Data Partitioning tool in H.264/AVC standard enables an encoder to rearrange the coded video data into three data partitions per slice in order to reduce the impact of transmission errors. These three data partitions are:

- DP-A = contains the header information including coding-mode, quantization parameters (QP), MVs, etc.
- DP-B = contains DC and AC coefficients for I-MBs.
- DP-C = contains DC and AC coefficients for P-MBs.

Specifically, DP-A is considered to be the most important for proper decoding of the video packet. Without DP-A, the coded video data conveyed by both DP-B and DP-C are seemingly unusable. In contrast, if DP-A is received, it is still possible for the decoder to make a reasonable attempt at reconstructing the video frame, even though the other partitions are missing due to channel errors.

We assume that, with proper interleaving, the packet loss for each data-partition can be modelled by an i.i.d. Bernoulli process. Let P_X denote the packet loss probability of DP-X where X = A, B or C. Based on this channel model, the channel-induced distortion for I-MBs is:

$$^I D_{c,n} = P_A D_{L,n} + (1 - P_A)(1 - P_B)D_{IR,n}$$
$$+ (1 - P_A)P_B D_{Icc,n} \qquad (5)$$

where $D_{L,n}$ ('L' for 'lost') is the distortion due to the lost of DP-A, $D_{IR,n}$ ('R' for 'received') is the distortion when both DP-A and DP-B are received, and lastly $D_{Icc,n}$ is the distortion when DP-A is received but DP-B is lost. We will show how each of these terms can be estimated.

We begin with $D_{L,n}$ which is the distortion due to the lost of DP-A. The loss of DP-A leads to the loss of the whole slice (or frame based on our assumptions) which means the decoder will have to conceal this lost in every MB in the slice by applying either frame-copy concealment or temporal concealment [8]. Let $cc(i)$ denote the matching pixel according to the occupied concealment algorithm so $\hat{f}_{n-1}^{cc(i)}$ and $\tilde{f}_{n-1}^{cc(i)}$ are the pixel values at pixel i in frame n after concealed at encoder and decoder. The distortion $D_{L,n}$ can be derived as follow:

$$D_{L,n} = E\left\{(\hat{f}_n^i - \tilde{f}_{n-1}^{cc(i)})^2\right\}$$
$$= E\left\{(\hat{f}_n^i - \hat{f}_{n-1}^{cc(i)} + \hat{f}_{n-1}^{cc(i)} - \tilde{f}_{n-1}^{cc(i)})^2\right\}$$
$$= E\left\{(\hat{f}_n^i - \hat{f}_{n-1}^{cc(i)})^2\right\} + E\left\{(\hat{f}_{n-1}^{cc(i)} - \tilde{f}_{n-1}^{cc(i)})^2\right\}$$
$$+ 2E\left\{(\hat{f}_n^i - \hat{f}_{n-1}^{cc(i)})(\hat{f}_{n-1}^{cc(i)} - \tilde{f}_{n-1}^{cc(i)})\right\}$$
$$= D_{ECP,n} + D_{c,n-1}$$
$$+ 2E\left\{(\hat{f}_n^i - \hat{f}_{n-1}^{cc(i)})(\hat{f}_{n-1}^{cc(i)} - \tilde{f}_{n-1}^{cc(i)})\right\} \qquad (6)$$

where

$$E\left\{(\hat{f}_n^i - \hat{f}_{n-1}^{cc(i)})^2\right\} = D_{ECP,n} \qquad (7)$$

$$E\left\{(\hat{f}_{n-1}^{cc(i)} - \tilde{f}_{n-1}^{cc(i)})^2\right\} \approx E\left\{(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i)^2\right\}$$
$$= D_{c,n-1} \qquad (8)$$

This derivation holds if the decoder uses frame-copy concealment where $cc(i) = i$. This means that we can use Eq. (6) to recursively compute $D_{L,n}$ by using $D_{c,n-1} = {}^I D_{c,n-1}$ and the initial condition

$$D_{c,0} = 0 \qquad (9)$$

which holds due to our assumption that the IDR frame is correctly received.

Note that the first two terms in Eq. (6) can be estimated at the encoder. The term $E\left\{(\hat{f}_n^i - \hat{f}_{n-1}^{cc(i)})(\hat{f}_{n-1}^{cc(i)} - \tilde{f}_{n-1}^{cc(i)})\right\}$ in Eq. (6) is what we called error-concealment cross-correlation. Since it will be easier to understand how this correlation term can be computed after knowing how $D_{IR,n}$ and $D_{Icc,n}$ can be estimated, we will first discuss how these two terms can be computed.

For the computation of $D_{IR,n}$, note that generally, I-MBs will be coded using intra-prediction, which, by definition, allows an I-MB to be predicted from its neighboring coded blocks in the same slice or beyond (unconstrained intra-prediction) thus utilizing spatial redundancy and increasing coding efficiency. Although it exploits predictive coding, I-MBs do not rely on the information beyond its own slice boundary unless unconstrained intra-prediction is employed so the I-MBs are distorted by neither spatial nor temporal error propagation. Moreover, they can stop further drift resulted from the occurrence of channel errors. Due to these properties and characteristics of I-MBs, it is not difficult to see that

$$D_{IR,n} = 0 \qquad (10)$$

We will now discuss the estimation of $D_{Icc,n}$. Since the JM 13.0 software, the H.264/AVC reference software, do not provide an appropriate mean to conceal the I-MB when DP-B is lost and it is not the purpose of this paper to propose new error concealment algorithm, we will use the default concealment method in JM 13.0 which sets all the pixel values in the lost I-MBs to 128. Hence the distortion $D_{Icc,n}$ is

$$D_{Icc,n} = E\left\{(\hat{f}_n^i - 128)^2\right\} \qquad (11)$$

Obviously, the distortion arises from this poor concealment method tend to be much higher than simple frame-copy concealment as can be seen from the simulation results in experimentation section.

In order to compute the error-concealment cross-correlation appeared in Eq. (6), we assume that $\tilde{f}_{n-1}^{cc(i)}$ and $(\hat{f}_n^i - \hat{f}_{n-1}^{cc(i)})$ are uncorrelated. We believed this assumption is reasonable as justified by the accuracy of the proposed model, see Section III. This term can be approximated as follows

$$E\left\{(\hat{f}_n^i - \hat{f}_{n-1}^{cc(i)})(\hat{f}_{n-1}^{cc(i)} - \tilde{f}_{n-1}^{cc(i)})\right\}$$
$$= E\left\{\hat{f}_{n-1}^{cc(i)}(\hat{f}_n^i - \hat{f}_{n-1}^{cc(i)})\right\} - E\left\{\tilde{f}_{n-1}^{cc(i)}(\hat{f}_n^i - \hat{f}_{n-1}^{cc(i)})\right\} \quad (12)$$
$$= E\left\{\hat{f}_{n-1}^{cc(i)}(\hat{f}_n^i - \hat{f}_{n-1}^{cc(i)})\right\} - E\left\{\tilde{f}_{n-1}^{cc(i)}\right\} E\left\{(\hat{f}_n^i - \hat{f}_{n-1}^{cc(i)})\right\}$$
$$\qquad (13)$$

As shown in Eq. (13), only the first moment of decoded pixel is required to calculate the cross-correlation term as the rest of the term is accessible at the encoder side. To develop the recursive formula to calculate the first moment of decoded pixel for I-MB, each combination of P_A and P_B must be statistically approximated. As described previously, I-MBs can be either completely lost or completely received and the concealed MBs are set to 128 for all pixels in the MBs if only DP-A is received. Otherwise, frame-copy concealment is

applied. Based on Bernoulli probability, the first moment of $\tilde{f}_n^i$ for I-MB can be represented as

$$E\left\{\tilde{f}_n^i\right\} = (1 - P_A)(1 - P_B) \times 0$$
$$+ (1 - P_A)P_B \times 128 + P_A E\left\{\tilde{f}_{n-1}^{cc(i)}\right\} \quad (14)$$

According to Eq. (9), no channel distortion on IDR frame so the reconstructed pixels at encoder and decoder must be the same and no concealment algorithm is applied.

$$E\left\{\tilde{f}_0^{cc(i)}\right\} = E\left\{\tilde{f}_0^i\right\} = E\left\{\hat{f}_0^i\right\} \quad (15)$$

Note that if correlation coefficient of $\hat{f}_n^i$ and $\hat{f}_{n-1}^{cc(i)}$ is relatively high i.e. higher than the threshold $thres_{cc}$ then we can deduce that $\hat{f}_n^i \approx \hat{f}_{n-1}^{cc(i)}$ as the pixels from both frames are very similar and thus $E\left\{(\hat{f}_n^i - \hat{f}_{n-1}^{cc(i)})(\hat{f}_{n-1}^{cc(i)} - \tilde{f}_{n-1}^{cc(i)})\right\} \approx 0$.

By substituting Eq. (6), (10) and (11) into (5), we get the complete distortion estimation model for I-MB.

C. P-MB Distortion

Basically, the pixels in decoded P-MBs are determined from the residuals (r_n^i) conveyed by DP-C and the estimated pixels in the reference frame pointed by the MVs ($f_{ref}^{mv(i)}$) in DP-A. We assume that only the previous frame ($n - 1$) will be used as the reference frame for motion compensation. The reconstructed pixel of the P-MB is defined by

$$f_n^i = r_n^i + f_{n-1}^{mv(i)} \quad (16)$$

Similar to intra-coding distortion, our proposed inter-coding distortion also applies Bernoulli process to estimate the expected channel-induced distortion for P-MBs. The estimation model for P-MBs can be written as

$$^P D_{c,n} = P_A D_{L,n} + (1 - P_A)(1 - P_C) D_{PR,n}$$
$$+ (1 - P_A) P_C D_{Pcc,n} \quad (17)$$

where

$$D_{L,n} = \text{ distortion resulted from lost-MB}$$
$$D_{PR,n} = \text{ distortion of receiving P-MB}$$
$$D_{Pcc,n} = \text{ distortion resulted from lost-residual}$$

The probability of DP-A and DP-C successfully reaching the receiver is $(1 - P_A)(1 - P_C)$ and the distortion of this case can be calculated from the mismatch of motion compensation. Let γ be the motion coverage i.e. the number of pixels being referenced by MVs over the total pixels in the frame. Thus $1 - \gamma$ is the ratio of unreferenced-pixels. $\gamma = 1.0$ implies that all pixels in the reference frame have been used and thus $E\left\{(\hat{f}_{n-1}^{mv(i)} - \tilde{f}_{n-1}^{mv(i)})^2\right\} = E\left\{(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i)^2\right\}$. Statistically, if γ is higher than a constant threshold $thres_{mv}$ then we can confidently represent $D_{PR,n}$ as distortion of previous frame weighted by the motion coverage γ. On the other hand, if γ is lower than $thres_{mv}$ then the weight is redefined as $\gamma + \gamma(1 - \gamma) = 2\gamma - \gamma^2$. As some of pixels in the reference frame can be referenced more than once, for low

γ some pixels in motion coverage area (γ) are referenced by $1 - \gamma$ times. This yields the distortion of receiving P-MB as

$$D_{PR,n} = E\left\{(\hat{f}_{n-1}^{mv(i)} - \tilde{f}_{n-1}^{mv(i)})^2\right\}$$
$$\approx m E\left\{(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i)^2\right\} = m D_{c,n-1} \quad (18)$$

where
$$m = \begin{cases} \gamma & \text{if } \gamma \geq thres_{mv} \\ (2\gamma - \gamma^2) & \text{otherwise} \end{cases} \quad (19)$$

In the case of losing DP-C, so no residue signal is available on the decoder. However, the frame still be able to use MVs in DP-A to predict the pixels at the current MB.

$$D_{Pcc,n} = E\left\{(\hat{f}_n^i - \tilde{f}_{n-1}^{mv(i)})^2\right\} \quad (20)$$

Substituting Eq. (16) into (20)

$$D_{Pcc,n} = E\left\{(\hat{f}_{n-1}^{mv(i)} + \hat{r}_n^i - \tilde{f}_{n-1}^{mv(i)})^2\right\}$$
$$= E\left\{(\hat{f}_{n-1}^{mv(i)} - \tilde{f}_{n-1}^{mv(i)})^2\right\} + E\left\{(\hat{r}_n^i)^2\right\}$$
$$+ 2E\left\{\hat{r}_n^i(\hat{f}_{n-1}^{mv(i)} - \tilde{f}_{n-1}^{mv(i)})\right\} \quad (21)$$

Assuming that $\hat{r}_n^i$ and $(\hat{f}_{n-1}^{mv(i)} - \tilde{f}_{n-1}^{mv(i)})$ is uncorrelated.

$$D_{Pcc,n} = m D_{c,n-1} + D_{r,n} \quad (22)$$

Although $D_{L,n}$ can be derived in the same method as that in intra-coding distortion, the calculation of error-propagated cross-correlation term is becoming more complex if the pixels belong to a P-MB. If the MV is lost with probability P_A, the decoder will conceal the P-MBs in the slice based on concealment algorithms being employed regardless of receiving residuals (DP-C) or not. If DP-A is correctly received, MVs will be used to predict pixel i in the current frame with pixel mv(i) in the reference frame. The prediction error will be added if DP-C is received. Hence the first moment of $\tilde{f}_n^i$ for P-MB is derived as

$$E\left\{\tilde{f}_n^i\right\} = (1 - P_A)(1 - P_C)\left(E\left\{\tilde{f}_{n-1}^{mv(i)}\right\} + E\left\{\hat{r}_n^i\right\}\right)$$
$$+ (1 - P_A)P_C E\left\{\tilde{f}_{n-1}^{mv(i)}\right\} + P_A E\left\{\tilde{f}_{n-1}^{cc(i)}\right\} \quad (23)$$

Where $E\left\{\tilde{f}_{n-1}^{mv(i)}\right\} = m E\left\{\tilde{f}_{n-1}^i\right\}$. To obtain the complete recursive formula for P-MB distortion, we substitute Eq. (6), (18) and (22) into (17).

III. EXPERIMENTAL EVALUATION

In order to verify our proposed frame-recursive distortion model, we carried out extensive simulations. The simulation environment is as follows. The source video is the QCIF standard test sequence *News* and is selected due to its rapid changes of spatial information and moderate temporal characteristics. 255 frames of the tested video sequence are encoded at 30.0 fps using JM13.0 the reference software codec that implementing the H.264/AVC video coding standard with various encoding configurations in the Extended profile. The encoder

encodes one slice per frame and uses fixed quantization parameter of 28 for both I-MBs and P-MBs. Constrained intra-prediction for all I-MBs and full-pel motion estimation with search range of 16 pels for all P-MBs. Fixed or variable block size coding can be used in all types of MB but deblocking filter is disabled as it not yet included in our model. Only frame-copy error concealment is used at the decoder. 500 simulations with random packet loss are collected for each given packet loss rate. The experimental results are then averaged and compared with the proposed model. Six different total packet loss probabilities (1%, 3%, 5%, 10%, 15% and 20%) were used in the simulations.

In order to accurately estimate the channel-induced distortion $D_{c,n}$, we need to model and analyze the packet loss behavior of the wireless channel. For a given total packet loss probability P, let N_A, N_B and N_C denote the number of packets for DP-A, B and C respectively, then the total packet loss probability is given by

$$P = \frac{P_A N_A + P_B N_B + P_C N_C}{N_A + N_B + N_C} \tag{24}$$

From this equation, we can derive the packet loss probability for I-frames when $\beta = 1$ by setting $N_C = 0$ and $N_A = N_B$, because I-frames consist of only I-MBs and DP-A and DP-B packets are produced for each I-frame thus $P = \frac{P_A + P_B}{2}$. In this case, if we consider prioritized transmission strategy, we can for example increase the retransmission limits of DP-A in order to decrease P_A. Note that due to the constant total packet loss rate P assumption, this will lead to an increase in the packet loss rate P_B for DP-B. This example shows that our work can be used to determine the optimal trade-off of the packet loss rate in bitstreams and channel distortion. In the first set of experiment, we let $P_A = P_B$ to examine the accuracy of our estimation model. Our simulation compares two distortion estimation models. The first model (*model1*) neglects the error-concealment cross-correlation term in Eq. (6) and the second model (*model2*), which is our proposed model, takes the cross-correlation term into account. Fig. 1 shows the performance of these two models when only I-frames ($\beta = 1$) are used for a total packet loss rate of 5%.
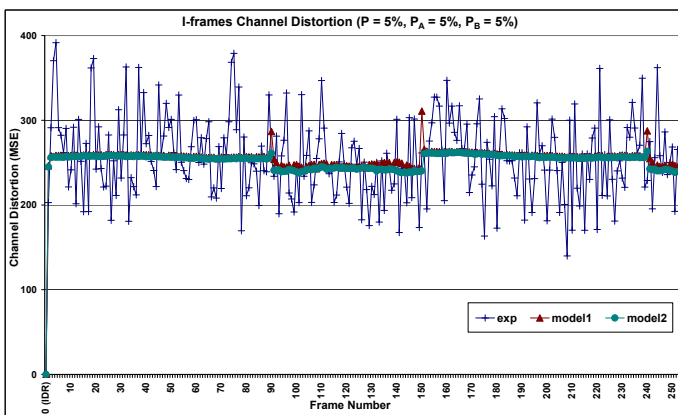
Not surprisingly, the experimental results in Fig. 1 show that cross-correlation term has minor effects in estimating I-frames distortion. As shown in Eq. (6), the cross-correlation term has the mathematical form of the multiplication of the current frame pixel difference of encoder reconstruction and concealment and that of previous frame implies that the cross-correlation is high when both error propagation and previous frame distortion are high. Therefore, if one of these is small then cross-correlation term is insignificant comparing to overall channel distortion. Due to typical characteristics of I-MBs that is able to suppress the frame-to-frame error-propagated distortion, the cross-correlation term can be neglected.

In our next set of experiments, we set $\beta = 0$, i.e. there is no I-MBs frames are used. When channel error distortion is estimated, only distortion resulted from P-MBs $^P D_{c,n}$ is considered. Since there are no DP-B packets, the packet loss rate can be described $P = \frac{P_A + P_C}{2}$. Fig. 2-3 shows the prediction performance for (*model1*) and (*model2*) for total packet loss rates of 5% and 10% respectively. Although the estimation model that ignores error-concealment cross-correlation (*model1*) mostly gives good distortion prediction, it performs poorly when rapid background change occurs (near frame 150). However, our proposed model *model2* gives good prediction for all frames.
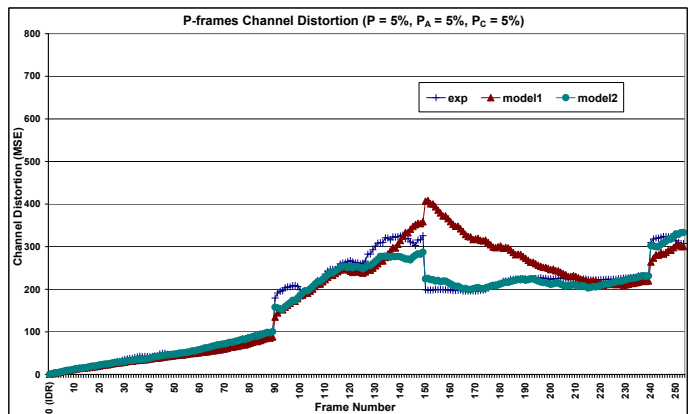

Fig. 2. Channel distortion of P-frames with $P = 5\%$, $\beta = 0\%$

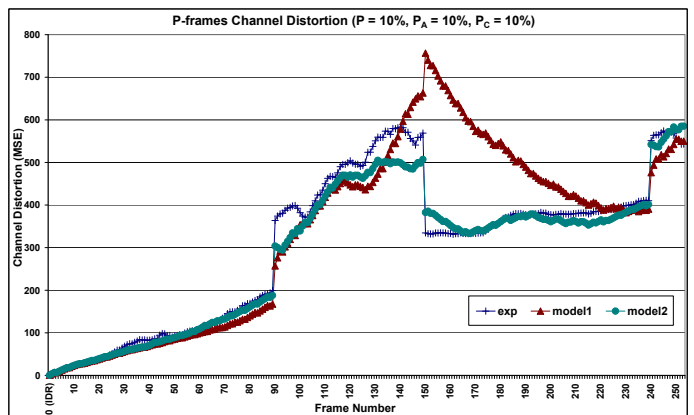

Fig. 1. Channel distortion of I-frames with $P = 5\%$

Fig. 3. Channel distortion of P-frames with $P = 10\%$, $\beta = 0\%$

In the next set of experiments, we enabled random intra macroblock refresh which means some MBs in P-slice will be intra-coded. As there are 3 partitions or packets per frame, it implies that $N_A = N_B = N_C$ and the packet loss rate in this case is $P = \frac{P_A + P_B + P_C}{3}$. We chose $P_A = 5\%$, $P_B = 5\%$ and $P_C = 5\%$ in our simulation, which results in $P = 5\%$. Two intra-rates (β), of 5% and 15%, are used to encode the video sequence. Fig. 4-5 illustrate the performance of our P-frames distortion model with intra-update enabled and also confirm our hypothesis that error-propagated cross-correlation can be suppressed by I-MBs as increasing intra-rate can in-turn reducing the inaccuracy of *model1*.
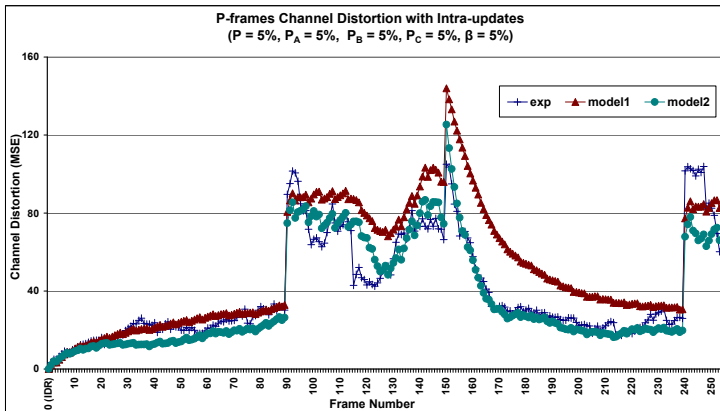

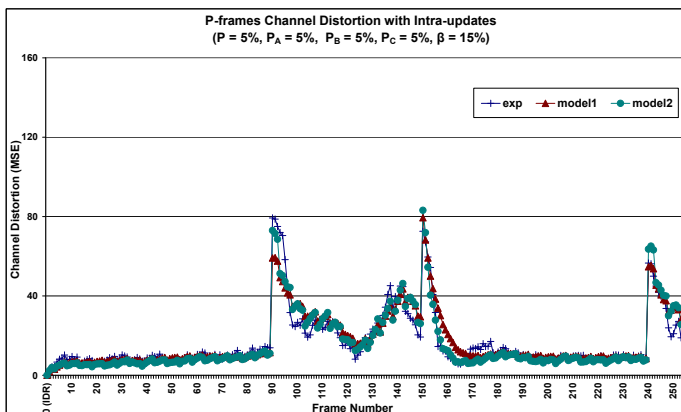Fig. 4.   Channel distortion of P-frames with $P = 5\%$, $\beta = 5\%$

Fig. 5.   Channel distortion of P-frames with $P = 5\%$, $\beta = 15\%$

TABLE I
SUMMARY OF THRESHOLDS IN DISTORTION ESTIMATION MODEL

News Sequence (255 frames @ 30fps)							
Packet Loss Prob. (%)					model1	model2	
P	P_A	P_B	P_C	β	$thres_{mv}$	$thres_{mv}$	$thres_{corr}$
5	5	-	5	0	0.972	0.976	0.975
10	10	-	10	0	0.971	0.975	0.978
5	5	5	5	5	0.950	0.926	0.830
5	5	5	5	15	0.990	0.830	0.821

In Table I, we summarize the motion coverage threshold $thres_{mv}$ and correlation coefficient threshold $thres_{corr}$ settings for different combinations of packet loss rate and intra-ratio presented here.

IV. CONCLUSION

We have combined the block-based distortion with frame-recursive model and derived a statistical channel-error distortion estimation model for wireless video transmission. The proposed work is a hybrid algorithm which operates much faster than pixel and block-recursive model but yet able to estimate the channel-loss distortion precisely by computing the error propagation at block level. It is also superior than the previous work in several aspects. Firstly, the model does not require the sample output distortion from decoder via feedback channel in order to apply least square fitting to compute the decoded video distortion. Secondly, we incorporate the cross-correlation approximation in our algorithm allowing the model to capture the distortion resulted from instantly changed scene. Thirdly, the model also takes into account both data partitioning and the newly introduced feature in H.264/AVC video coding standard namely variable block size motion compensation. Finally, the distortion estimation algorithm can be applied to RD model where prioritized transmission strategy is considered e.g. reducing the packet loss rate in one bitstream while increasing other bitstreams' loss rate. The accuracy of the proposed channel-induced distortion model is observed through the comparison with the simulation results. Although our cross-correlation model (*model2*) and the model without cross-correlation (*model1*) yields the same results in I-frames, our model (*model2*) does perform better in P-frames where error propagation has significant impact on overall distortion.

REFERENCES

[1] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra mode switching for packet loss resilience," *IEEE J. Sel. Areas Commun.*, vol. 18, pp. 966–976, Jun. 2000.

[2] H. Yang and K. Rose, "Recursive end-to-end distortion estimation with model-based cross-correlation approximation," in *Proc. IEEE Conf. Image Processing*, Barcelona, Spain, 2003, pp. 469–472.

[3] K. Stuhlmueller, N. Faerber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Sel. Areas Commun.*, vol. 18, pp. 1012–1032, Jun. 2000.

[4] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks," *IEEE J. Sel. Areas Commun.*, vol. 18, pp. 952–965, Jun. 2000.

[5] S. Ekmekci and T. Sikora, "Recursive decoder distortion estimation based on ar(1) source modeling for video," in *Proc. IEEE Conf. Image Processing*, Singapore, 2004, pp. 187–190.

[6] Y. Wang, Z. Wu, and J. M. Boyce, "Modeling of transmission-loss-induced distortion in decoded video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, pp. 716–732, Jun. 2006.

[7] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/avc in wireless environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 657–673, Jul. 2003.

[8] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 560–576, Jul. 2003.

[9] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 511–523, Jun. 2002.