Modeling of Transmission-Loss-Induced Distortion in Decoded Video

Yao Wang, *Fellow, IEEE*, Zhenyu Wu, *Member, IEEE*, and Jill M. Boyce, *Member, IEEE*

***Abstract*—This paper analyzes the distortion in decoded video caused by random packet losses in the underlying transmission network. A recursion model is derived that relates the average channel-induced distortion in successive P-frames. The model is applicable to all video encoders using the block-based motion-compensated prediction framework (including the H.261/263/264 and MPEG1/2/4 video coding standards) and allows for any motion-compensated temporal concealment method at the decoder. The model explicitly considers the interpolation operation invoked for motion-compensated temporal prediction and concealment with sub-pel motion vectors. The model also takes into account the two new features of the H.264/AVC standard, namely intraprediction and inloop deblocking filtering. A comparison with simulation data shows that the model is very accurate over a large range of packet loss rates and encoder intrablock rates. The model is further adapted to characterize the channel distortion in subsequent received frames after a single lost frame. This allows one to easily evaluate the impact of a single frame loss.**

***Index Terms*—Deblocking filter, end-to-end distortion, error concealment, error propagation, H.264/AVC, intraprediction, packet loss.**

I. Introduction

THE QUALITY of the decoded video in a networked video application depends both on the quantization incurred at the encoder and the distortion due to channel errors (e.g., packet losses) that have occurred during transmission. We refer the latter as the channel distortion, and it depends on channel loss characteristics, the coder error resilience features (e.g., the intrarate, which is the frequency at which a macroblock (MB) is coded in the INTRA mode), and the decoder error concealment method. Accurate modeling of the channel-induced distortion is important for jointly determining parameters for source coding and channel error control, and for rate-distortion-optimized mode decision in the encoder.

The past research for determining the channel (or total) distortion can be categorized based on whether the distortion is estimated and tracked at the pixel level, at the MB level, or at the frame level. The well-known recursive optimal per-pixel estimation (ROPE) method [1] recursively calculates the first- and second-order moments of the decoded value for each pixel, from which one can determine the total distortion [in terms of the mean squared error (MSE)] of each pixel. A problem with the ROPE method is that it is only applicable when the motion vectors (MVs) have integer accuracy and when the decoder conceals lost MBs using the simple frame copy method. Although the ROPE method provides very accurate distortion estimation under the above constraints, it requires intense computation, since it involves tracking two moments at every pixel. The extension of ROPE to consider sub-pel motion compensation is considered in [2], which requires substantially more computation.

The work in [3] computes and tracks the average distortion at each MB. In our opinion, the recursion formula used to determine the distortion caused by predicting from previously concealed blocks does not account for the error propagation effect precisely. It also ignores the error propagation associated with temporal error concealment. As with the ROPE method, it does not consider sub-pel motion compensation. The more recent work in [4] also estimates the total distortion at the MB level. It determines the distortion of the current MB by considering the cases when the corresponding matching block in the previous frame is received and lost, separately. Thus, it needs to track the distortion when a block is received and lost separately, for each MB in the past frame. Finally, it assumes that encoder motion vectors are available at the decoder for error concealment. It does attempt to capture the effect of filtering associated with sub-pel motion compensation and deblocking through an attenuation factor, which was determined manually for each test sequence.

The proceeding pixel-level and MB-level distortion estimation methods can be used for encoder mode decision. They can also be used to determine the distortion at each pixel or the average distortion of each MB or each frame, given the chosen coding modes and MVs for all MBs. However, these methods cannot estimate the expected distortion before encoding a sequence (and, hence, deciding on the coding mode and MV of each MB), based on a given intrarate. In many applications, it is desirable to estimate the average distortion for an entire video sequence or a segment using different quantization parameters and intrarates, without coding the sequence with all possible settings. Such estimation is required either for determining the video coder configuration (e.g., the quantization parameter and the intrarate) to meet a certain quality and/or rate target, or for jointly optimizing the video coder configuration and the channel code rate, under a total channel bandwidth constraint.

To enable the aforementioned operations, it is desirable to have a mathematical model that relates the average channel (or total) distortion over individual frames or a group of frames with the average intrarate and packet-loss rate. Such a model was presented in [5], which was derived by characterizing the spatial-temporal error propagation behavior after an initial loss using a linear filter, with a leakage parameter accounting for the

Manuscript received October 24, 2005. A preliminary version of this work was presented at IEEE International Conference on Multimedia and Expo, June 2005. This paper was recommended by Associate Editor H. Gharavi.

Y. Wang is with the Polytechnic University, Brooklyn, NY 11201 USA (e-mail: yao@poly.edu).

Z. Wu and J. M. Boyce are with Corporate Research, Thomson, Inc., Princeton, NJ 08540 USA (e-mail: zhenyu.wu@thomson.net; jill.boyce@thomson.net).

Digital Object Identifier 10.1109/TCSVT.2006.875203

filtering incurred for sub-pel motion compensation and/or deblocking. It further assumes that the effects of losses occurring at different times are additive, which is only valid at low-loss rates. There have been concerns that its estimation accuracy may be inadequate [4]. He *et al.* [6] developed a frame-level recursion formula, which relates the channel-induced distortion in a current frame with that in the previous frame. However, it does not take into account of the filtering incurred for sub-pel motion compensation and/or deblocking filtering, and assumes the decoder uses frame copy for error concealment.

All the prior work, at the pixel, MB or frame level, does not consider intraprediction. Nor do they consider deblocking filtering explicitly. Some of them also assume that the decoder uses the simple frame copy method for error concealment. Intraprediction and deblocking filtering are two new features of the latest H.264/AVC video-coding standard [7]. Furthermore, motion-compensated temporal concealment is known to improve the decoded video quality significantly. Therefore, it is important to consider these features in channel distortion modeling. The work presented here is similar to [6] in that we also model the average frame-level distortion through a recursion formula. We, however, extend the scope of that work substantially. We consider both inter and intraprediction, and deblocking. We also allow for motion-compensated temporal concealment at the decoder. For motion-compensated prediction and concealment, we explicitly consider the interpolation operation incurred when motion vectors are noninteger. Finally, we also adapt the proposed model to characterize the channel distortion in subsequent received frames after a single lost frame. This allows one to easily evaluate the impact of a single frame loss. This can be helpful, e.g., for performing unequal error protection at the frame level.

This paper is organized as follows. Section II introduces the notations and assumptions underlying our model. Section II develops a frame-level recursion formula for the channel-induced distortion due to random packet loss, starting from the case when the encoder does not employ intraprediction and deblocking filtering, and then moving on to consider intraprediction and deblocking filtering, successively. Section IV verifies the proposed model with simulation data. Section V describes the proposed model for channel distortion due to a single frame loss, and examines its accuracy using simulation data. Section VI concludes the paper by summarizing the main results, and discussing limitations of the current work and possible solutions.

II. Notation and General Assumption

A. Definition of Frame-Level Channel Distortion

Let f_n^i denote the original pixel value in frame n and pixel i, $\hat{f}_n^i$ the reconstructed signal at the encoder, and $\tilde{f}_n^i$ the reconstructed signal at the decoder. In the presence of transmission errors, $\hat{f}_n^i$ and $\tilde{f}_n^i$ are generally different. The total distortion in terms of MSE at pixel i is defined as $D_n^i = E_c\{(f_n^i - \tilde{f}_n^i)^2\}$, where $E_c\{\cdot\}$ represents the expectation taken over all possible channel realizations. The average distortion over all pixels in frame n is

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

where $E_a\{\cdot\}$ denotes the averaging-over-pixel operation. For the remainder of this paper, we use $E\{\cdot\}$ to denote the concatenated operation $E_a\{E_c\{\cdot\}\}$.

The total average distortion can be written as

$$\begin{aligned} D_n &= E\left\{\left(f_n^i - \hat{f}_n^i + \hat{f}_n^i - \tilde{f}_n^i\right)^2\right\} \\ &= E_a\left\{\left(f_n^i - \hat{f}_n^i\right)^2\right\} + E\left\{\left(\hat{f}_n^i - \tilde{f}_n^i\right)^2\right\} \\ &\quad + 2E\left\{\left(f_n^i - \hat{f}_n^i\right)\left(\hat{f}_n^i - \tilde{f}_n^i\right)\right\}. \end{aligned}$$

If we assume the encoder-introduced error $f_n^i - \hat{f}_n^i$ (caused by quantization) and the channel-induced error $\hat{f}_n^i - \tilde{f}_n^i$ are uncorrelated, then the total distortion can be decomposed as

$$D_n = D_{e,n} + D_{c,n} \tag{2}$$

where

$$D_{e,n} = E_a\left\{\left(f_n^i - \hat{f}_n^i\right)^2\right\} \tag{3}$$

is the average encoder-induced distortion per pixel in frame n, and

$$D_{c,n} = E\left\{\left(\hat{f}_n^i - \tilde{f}_n^i\right)^2\right\} \tag{4}$$

is the average channel-induced distortion per pixel in frame n. He *et al.* have shown with experimental data [6] that the encoder-introduced error and the channel-induced error are indeed quite uncorrelated. Since the encoder distortion can be determined accurately at the encoder, the challenging problem in determining the total distortion lies in the estimation of the channel distortion. This paper deals with the modeling of the average channel-induced distortion in each frame $D_{c,n}$.

B. Assumptions Underlying the Model

Encoder: We assume a video sequence is partitioned into groups of pictures (GoPs) and each GoP starts with an I-frame, followed by P-frames.[1] In an I-frame, every MB is coded in the INTRA mode, without making use of previous frame information. In the INTRA mode in the earlier video-coding standards, pixels are coded directly using transform coding. In the H.264 standard, intraprediction is applied, by which an MB is first predicted using previously coded neighboring pixels in this frame, and then the prediction error is coded using transform coding. In a P-frame, an MB is coded either in the INTRA mode or INTER mode, by which the MB is predicted by a corresponding MB in a previous frame. When the motion vector (MV) for this MB is an integer vector, pixels in this MB are predicted by corresponding pixels in the best-matching MB in the previous frame directly. But when the MV is noninteger, each pixel in the current MB is predicted by interpolating from a set of corresponding pixels in the previous frame. We will refer to INTRA-coded MBs as I-MBs, and INTER-coded MBs as P-MBs. We assume the length

[1]Although we do not consider B-frames in this paper, the proposed model can be used to model channel distortions in successive P-frames, even if there are B-frames between two adjacent P-frames. The model can also be extended to consider the channel distortion in B-frames.

of each GoP is N and the percentage of I-MBs in each P-frame is β_n. An MB can be coded in the INTRA mode either for coding efficiency or for error resilience. Generally, the number of I-MBs due to coding efficiency varies from frame to frame. Therefore, β_n may not be a constant.

Packetization and Loss Pattern: We assume that the MBs in a frame are grouped into slices so that there is no dependency between coded data in separate slices, and each slice has its own header, and is carried in a separate transport packet. We also assume that the loss of any bits in a packet will make the entire corresponding video slice undecodable. We further assume that with proper packet interleaving, the packet (and, hence, slice) loss event can be characterized as an i.i.d. random process with an average loss rate P. We do not make any specific assumptions on the slice length (in terms of the number of MBs contained) and pattern (in terms of the order by which the MBs are put into a slice). The slice length may be constant or variable. We only require that a slice contain either a subset or all MBs in a frame, but not MBs from different frames, and that all coded data (including MVs) for an MB are put into the same slice.[2] Note that the slice pattern can have a significant impact on the accuracy of error concealment.

Decoder: We assume the decoder performs error concealment on MBs that are contained in lost slices. Generally, one can use either spatial or temporal concealment, or both. Here, we consider only temporal concealment for P-frames, by which the MV of a missing MB is first estimated (typically based on the MVs of the adjacent MBs in the current frame or the previous frame), and then the MB is replaced by the corresponding MB in the previous frame based on the estimated MV. When the estimated MV is noninteger, each pixel is a weighted average of some pixels in the previous frame. A commonly used concealment method is frame copy, which is a special case discussed here with all of the MVs set to zero. We assume the I-frames are concealed by spatial concealment, where a missing pixel is recovered from a weighted average of nearby received or previously concealed pixels in the same frame. Because the I-frames are concealed by spatial interpolation, any error propagation stops at an I-frame. Therefore, we only need to model the channel error propagation in successive P-frames.

III. Channel Distortion Model for Random Packet Losses

A. Overview of the Model

The proposed channel distortion model is a frame-recursive model, relating the channel distortion of frame n with that of frame $n-1$ within the same GoP. It has the general form of

$$D_{c,n} = PD_{\mathrm{ECP},n} + \alpha(\beta_n, P)D_{c,n-1}, \quad n = 1, 2, \ldots, N-1. \quad (5)$$

The variable $D_{\mathrm{ECP},n}$ denotes the average concealment distortion of frame n, defined as the average channel distortion per pixel of this frame, if only one slice in this frame is lost and no other slices in this frame or previous frames are lost (to be defined more rigorously later). This is the new distortion introduced when any slice in frame n experiences packet losses and its value depends on the underlying error concealment method and the characteristics of that frame. We assume this term can be estimated at the encoder.

From the recursion formula in (5), $D_{c,n}$ is the sum of the concealment distortion in this frame $D_{\mathrm{ECP},n}$ and the product of the channel distortion in the previous frame $D_{c,n-1}$ by a factor α, which depends on the loss probability P and intrarate β_n. We will call α the error propagation factor, because it characterizes how the channel distortion in a previous frame propagates into the current frame. The particular form of $\alpha(\cdot)$ depends on the encoder configuration (with or without intraprediction and deblocking filtering, integer versus sub-pel motion estimation) and the decoder error-concealment algorithm (with or without motion compensation, with integer or sub-pel MV).

In the following subsections, we derive the forms of α for different encoder and decoder configurations. We will start with the simplest case when the encoder does not use intraprediction nor deblocking filtering, and for temporal prediction, uses only integer pel MVs. Similarly, the decoder uses only integer-pel MVs for error concealment. We will then move on to consider sub-pel MVs for encoder motion compensation and decoder error concealment, intraprediction, and deblocking filtering in that order. As will be seen, when constrained intraprediction is used, the expression for α stays the same as the case without intraprediction. However, nonconstrained intraprediction calls for a different form of α to account for the spatial error propagation caused by intraprediction.

In the subsequent derivation, we assume that if a slice is lost in frame n (with probability P), all of the MBs in this slice will be concealed using motion-compensated temporal concealment, with an average distortion $D_{\mathrm{L},n}$. If a slice is received (with probability $1-P$), the decoded MBs in this slice can still have channel distortion due to errors in the previous frame or previously coded pixels in the same frame. Denoting the average distortion in received I-MBs and P-MBs by $D_{\mathrm{IR},n}$ and $D_{\mathrm{PR},n}$, respectively, the average channel distortion is

$$D_{c,n} = (1-P)((1-\beta_n)D_{\mathrm{PR},n} + \beta_n D_{\mathrm{IR},n}) + PD_{\mathrm{L},n}. \quad (6)$$

This is the general relation we use to derive the distortion models for different cases.

B. Integer-Pel Motion Vectors, No Intraprediction, No Deblocking Filtering

For ease of understanding, we first develop the recursion model assuming the encoder does not use intraprediction or deblocking filtering. In this case, for a received I-MB, there will be no channel distortion (i.e., $D_{\mathrm{IR},n} = 0$). For a P-MB, even if it is received, its reconstruction may have channel distortion due to errors in the previous frame. Because we assume interprediction uses only integer MVs, each pixel f_n^i in this MB is predicted from a pixel $\hat{f}_{n-1}^{u(i)}$, where $u(i)$ refers to the spatial index of the matching pixel in frame $n-1$. In the receiver, if the MB is received, the prediction is based on $\tilde{f}_{n-1}^{u(i)}$. Because the prediction error signal for this MB is received correctly, the

[2]The proposed model will still work if the coding mode and motion information are transported separately and reliably. In that case, P denotes the loss rate for the slice containing texture data.

channel distortion is purely caused by the mismatch between the prediction references, with distortion

$$D_{\mathrm{PR},n} = E\left\{\left(\hat{f}_n^i - \tilde{f}_n^i\right)^2\right\} = E\left\{\left(\hat{f}_{n-1}^{u(i)} - \tilde{f}_{n-1}^{u(i)}\right)^2\right\} = D_{c,n-1}. \quad (7)$$

Note that, by definition, $E\{(\hat{f}_{n-1}^{u(i)} - \tilde{f}_{n-1}^{u(i)})^2\}$ is the average distortion among all pixel locations $u(i)$ corresponding to all possible i over a frame. In deriving (7), we have assumed that corresponding $u(i)$ for all possible i cover an entire frame evenly, so that the average over $u(i)$ is equal to the average of all pixels in frame $n-1$. Note that in certain special cases, such as during a camera zoom, $u(i)$ for all possible i in f_n may correspond to a subregion in f_{n-1}, making (7) less accurate.

If an MB is lost, we assume that it will be concealed using temporal concealment with an estimated MV, regardless of the coding mode. Let $s(i)$ represent the estimated matching pixel in frame $n-1$ for pixel i based on the estimated MV, then the concealed value for pixel i is $f_{\mathrm{ECP,integer\,MV}}^i = \tilde{f}_{n-1}^{s(i)}$. The channel distortion is thus

$$\begin{aligned} D_{\mathrm{L},n} &= E\left\{\left(\hat{f}_n^i - f_{\mathrm{ECP,integer\,MV}}^i\right)^2\right\} = E\left\{\left(\hat{f}_n^i - \tilde{f}_{n-1}^{s(i)}\right)^2\right\} \\ &= E\left\{\left(\hat{f}_n^i - \hat{f}_{n-1}^{s(i)} + \hat{f}_{n-1}^{s(i)} - \tilde{f}_{n-1}^{s(i)}\right)^2\right\} \\ &= E_a\left\{\left(\hat{f}_n^i - \hat{f}_{n-1}^{s(i)}\right)^2\right\} + E\left\{\left(\hat{f}_{n-1}^{s(i)} - \tilde{f}_{n-1}^{s(i)}\right)^2\right\} \\ &= D_{\mathrm{ECP},n} + D_{c,n-1} \end{aligned} \quad (8)$$

with

$$D_{\mathrm{ECP},n} = E\left\{\left(\hat{f}_n^i - \hat{f}_{n-1}^{s(i)}\right)^2\right\}. \quad (9)$$

Note that $\hat{f}_{n-1}^{s(i)}$ would have been the concealed value for f_n^i if there were no channel-induced distortion up to frame $n-1$. Therefore, $\hat{f}_n^i - \hat{f}_{n-1}^{s(i)}$ represents the newly introduced error at pixel i caused by the loss in frame n only. In deriving the above result, we have assumed that this new error is uncorrelated with the channel-induced error up to frame $n-1$ at the corresponding pixel $s(i)$, $\hat{f}_{n-1}^{s(i)} - \tilde{f}_{n-1}^{s(i)}$. We believe this is a quite reasonable assumption. Similar assumptions are used to derive subsequent results.

The term $D_{\mathrm{ECP},n}$ represents the distortion associated with a particular temporal concealment algorithm, in the absence of error propagation from previous frames. We will refer to it as concealment distortion. For example, if we use the simple frame-copy error concealment method, then $s(i) = i$ and $D_{\mathrm{ECP},n} = E\{(\hat{f}_n^i - \hat{f}_{n-1}^i)^2\}$ is the mean squared difference between two successively coded frames. Generally, the underlying slice pattern affects the estimation of the MVs and, hence, $D_{\mathrm{ECP},n}$. For example, if an entire frame is coded into a single slice, the MVs in a current frame have to be estimated from the MVs in the previous frame or simply set to zero. On the other hand, if a frame is divided into two interleaving slices, then the MVs in one slice can be more accurately estimated from the MVs for the neighboring MBs in the other slice. As is clear from the definition, the concealment distortion $D_{\mathrm{ECP},n}$ also depends on the encoder quantization parameter.

Substituting $D_{\mathrm{IR},n} = 0$ and (7) and (8) into (6), we obtain the recursion model in (5) with

$$\alpha = (1-P)(1-\beta_n) + P = 1 - \beta_n(1-P). \quad (10)$$

Note that in this case, α is equal to the percentage of pixels at which error propagation will continue. Generally, $0 < \alpha < 1$, except in the unlikely case of $\beta_n = 0$, or $P = 1$, or $\beta = 1$ and $P = 0$. Clearly, by increasing the intrarate, we can reduce α, so that the error propagation decays faster.

C. Sub-Pel Motion Vectors, No Intraprediction, No Deblocking Filtering

In all of the state-of-the-art video coders, sub-pel MVs are used to improve the prediction accuracy. To take into account the interpolation operation incurred for sub-pel MVs, we assume that a pixel f_n^i in a P-MB in frame n is predicted by a weighted sum of several neighboring pixels in frame $n-1$, denoted by

$$f_{p,p,e}^i = \sum_{l=1}^{L_{p,p}} a_l \hat{f}_{n-1}^{u_l(i)} \quad (11)$$

where $u_l(i)$ refers to the spatial index of the lth pixel in frame $n-1$ that was used to predict f_n^i. Note that this formulation is also applicable to overlapped block motion compensation (OBMC) employed in the H.263 codec. The interpolation coefficients a_l should satisfy $\sum_l a_l = 1$. Note that the values for $L_{p,p}$ and a_l depend on the MV for the MB, and the interpolation filter employed for sub-pel motion compensation. For example, with the bilinear interpolation filter, if the MV is half-pel in both horizontal and vertical direction, then $L_{p,p} = 4, a_l = 1/4$. On the other hand, if the MV is half-pel in one direction but integer in the other $L_{p,p} = 2, a_l = 1/2$. If the MV is integer in both directions, $L_{p,p} = 1, a_l = 1$.

At the receiver, if the MB is received correctly, the prediction is based on

$$f_{p,p,d}^i = \sum_{l=1}^{L_{p,p}} a_l \tilde{f}_{n-1}^{u_l(i)}. \quad (12)$$

Defining $e^i = \hat{f}_{n-1}^i - \tilde{f}_{n-1}^i$, the average channel distortion for the pixels associated with a particular set of $L_{p,p}, a_l$ is

$$\begin{aligned} &D_{\mathrm{PR},n,\mathrm{given}\ a_l} \\ &= E\left\{\left(f_{p,p,e}^i - f_{p,p,d}^i\right)^2\right\} \\ &= E\left\{\left(\sum_{l=1}^{L_{p,p}} a_l\left(\hat{f}_{n-1}^{u_l(i)} - \tilde{f}_{n-1}^{u_l(i)}\right)\right)^2\right\} \\ &= \sum_{l=1}^{L_{p,p}} a_l^2 E\left\{\left(e^{u_l(i)}\right)^2\right\} + \sum_{l,k,l\neq k} a_l a_k E\left\{e^{u_l(i)} e^{u_k(i)}\right\} \\ &= \left(\sum_{l=1}^{L_{p,p}} a_l^2 + \rho \sum_{l,k,l\neq k} a_l a_r\right) D_{c,n-1}. \end{aligned} \quad (13)$$

Note that generally, the channel induced errors on neighboring pixels are correlated, especially if these pixels belong to the same slice. To simplify the analysis, while deriving (13), we assumed that the correlation coefficients between errors in every two neighboring pixels are the same, represented by ρ. One can think of ρ as the average correlation coefficient of the channel-induced errors.

The fact that different values of $L_{p,p}, a_l$ are used for different received P-MBs can be accounted for by taking the average of the factors $\sum_{l=1}^{L_{p,p}} a_l^2 + \rho \sum_{l,k,l\neq k} a_l a_k$ used over all P-MBs in a frame, and denote the average value as

$$a = E_a \left\{ \sum_{l=1}^{L_{p,p}} a_l^2 + \rho \sum_{l,k,l\neq k} a_l a_k \right\}. \tag{14}$$

Further assuming that this average value is the same in different frames, the average channel distortion for received P-MBs with all possible MVs is

$$D_{\mathrm{PR,Model1},n} = aD_{c,n-1}. \tag{15}$$

The definition of a in (14) assumes that the correlation between channel-induced errors in neighboring pixels is a constant. Because the received I-MBs do not have any channel errors, the average correlation in a frame is likely to decrease with the intrarate in this frame. We have confirmed this conjecture with our simulation data, and found that a more accurate model is

$$D_{\mathrm{PR,Model2},n} = (a + b(1-\beta_{n-1}))D_{c,n-1} \tag{16}$$

with

$$\begin{aligned} a &= E_a \left\{ \sum_{l=1}^{L_{p,p}} a_l^2 \right\} \\ b &= \frac{\rho}{1-\beta_{n-1}} E_a \left\{ \sum_{l,k,l\neq k} a_l a_k \right\}. \end{aligned} \tag{17}$$

If an MB is lost, it will be concealed using temporal concealment with an estimated MV. Generally, the estimated MV may also be a sub-pel vector, and the concealed value can be denoted by

$$f_{\mathrm{ECP}}^i = \sum_{l=1}^{L_{c,p}} h_l \tilde{f}_{n-1}^{s_l(i)} \tag{18}$$

where $L_{c,p}$ represents the number of pixels used for temporal concealment, and $s_l(i)$ is the spatial index of the lth pixel in frame $n-1$ used for interpolating pixel i in frame n. Note that because the estimated MV may not be the same as the MV used in the encoder, $L_{c,p}$ and $s_l(i)$ generally differ from $L_{p,p}$ and $u_l(i)$ for the same pixel.

The channel distortion for MBs with the same set of h_l is

$$\begin{aligned} &D_{L,n,\text{given } h_l} \\ &= E\left\{ \left(\hat{f}_n^i - f_{\mathrm{ECP}}^i \right)^2 \right\} \\ &= E\left\{ \left(\hat{f}_n^i - \sum_{l=1}^{L_{c,p}} h_l \tilde{f}_{n-1}^{s_l(i)} \right)^2 \right\} \\ &= E\left\{ \left(\hat{f}_n^i - \sum_{l=1}^{L_{c,p}} h_l \hat{f}_{n-1}^{s_l(i)} + \sum_{l=1}^{L_{c,p}} h_l \hat{f}_{n-1}^{s_l(i)} - \sum_{l=1}^{L_{c,p}} h_l \tilde{f}_{n-1}^{s_l(i)} \right)^2 \right\} \\ &= E\left\{ \left(\hat{f}_n^i - \sum_{l=1}^{L_{c,p}} h_l \hat{f}_{n-1}^{s_l(i)} \right)^2 \right\} + \sum_{l,k} h_l h_k E\left\{ e^{s_l(i)} e^{s_k(i)} \right\} \\ &= D_{\mathrm{ECP},n} + \left(\sum_{l=1}^{L_{c,p}} h_l^2 + \rho \sum_{l,k,l\neq k} h_l h_k \right) D_{c,n-1} \end{aligned}$$

with

$$D_{\mathrm{ECP},n} = E\left\{ \left(\hat{f}_n^i - \sum_{l=1}^{L_{c,p}} h_l \hat{f}_{n-1}^{s_l(i)} \right)^2 \right\}. \tag{19}$$

Averaging over MBs with different h_l yields

$$D_{\mathrm{L},n} = D_{\mathrm{ECP},n} + hD_{c,n-1} \tag{20}$$

with

$$h = E_a \left\{ \sum_{l=1}^{L_{c,p}} h_l^2 + \rho \sum_{l,k,l\neq k} h_l h_k \right\}. \tag{21}$$

Note that $\sum_{l=1}^{L_{c,p}} h_l \hat{f}_{n-1}^{s_l(i)}$ would have been the concealed value using the same estimated MV in the absence of transmission errors in the previous frame. Therefore, $D_{\mathrm{ECP},n}$ represents the average concealment distortion for frame n when loss occurs only in this frame. When deriving the above results, we have again assumed that the concealment error at frame n is uncorrelated with the channel-induced errors in frame $n-1$. We also assumed that the channel-induced errors in neighboring pixels in frame $n-1$ have the same pairwise correlation coefficient ρ.

Substituting (15) and (20) and $D_{\mathrm{IR},n} = 0$ into (6) yields the recursion model in (5) with

$$\text{Model1}: \quad \alpha = a(1-\beta_n)(1-P) + hP. \tag{22}$$

With the more accurate expression for $D_{\mathrm{PR},n}$ in (16), the α factor is

$$\text{Model2}: \quad \alpha = (a+b(1-\beta_{n-1}))(1-\beta_n)(1-P)+hP. \quad (23)$$

We will refer to the distortion model using (5) and (22) as Model1, and the one using (5) and (23) as Model2. Model2 is more accurate than Model1, but it requires one extra parameter.

Comparing (22) to (10), we see that the effect of sub-pel MV is to introduce the constants a and h into the error propagation factor. Recall that the interpolation filter coefficients satisfy $|a_l| \leq 1$ and $\sum_l a_l = 1$. Because the correlation coefficient satisfies $|\rho| \leq 1$, we have $0 < a \leq 1$. In the special case of $\rho = 1$, we have $a = (\sum_l a_l)^2 = 1$. Similarly, $0 < h \leq 1$. Therefore, the error propagation factor α with sub-pel MV is typically smaller than when only integer-pel MVs are used. Thus, the spatial filtering incurred either for sub-pel motion-compensated prediction or sub-pel temporal concealment has the effect of attenuating the temporal error propagation, as is well known in the video-coding community.

To apply the above model for real video data, the model parameters a, b, and h can be predetermined by fitting the model to the measured distortion data from simulations. This is discussed further in Section IV-B. The concealment distortion $D_{\mathrm{ECP},n}$ can be directly measured by the encoder after coding frame n, by randomly setting certain slices in this frame as lost and running the assumed error concealment method on selected MBs in the lost slices.

The distortion model in [6] has the same form as (5), (22) but with $h = 1$, because it assumes frame copy for concealment. The constant a was introduced to account for the so-called motion randomness. The derivation here shows clearly the relation of a with the interpolation coefficients used for motion compensation with sub-pel motion vectors and the correlation between channel-induced errors in neighboring pixels. The model in [6] assumes the concealment distortion is proportional to the frame difference square $D_{\mathrm{ECP},n} = eE\{(f_n^i - f_{n-1}^i)^2\}$, where e is a model parameter. This assumption is only valid for the frame-copy error-concealment method. In our case, we assume $D_{\mathrm{ECP},n}$ can be measured by the encoder, by running the decoder error concealment method on selected MBs.

D. With Intraprediction, No Deblocking Filtering

When an I-MB makes use of intraprediction, there will be error propagation within the same frame. We now extend the previous derivation to consider this effect. We will assume that a pixel f_n^i in an I-MB in frame n is predicted by a weighted sum of several previously coded neighboring pixels in the same frame, denoted by

$$f_{p,i,e}^i = \sum_{l=1}^{L_{p,i}} c_l \hat{f}_n^{q_l(i)} \quad (24)$$

where $q_l(i)$ refers to the spatial index of the lth previously coded pixel in frame n that is used to predict f_n^i. For example, in the H.264 standard, there are many different modes of intraprediction, each leads to a different set of values for $L_{p,i}, q_l(i)$ and c_l. In each case, the coefficients c_l satisfy $\sum_l c_l = 1$.

If an I-MB is received, the intrapredicted value at the decoder will be

$$f_{p,i,d}^i = \sum_{l=1}^{L_{p,i}} c_l \tilde{f}_n^{q_l(i)}. \quad (25)$$

The channel distortion will be due to the difference in the predicted value only. Generally, the neighboring pixels used to predict a current pixel may come from either I-MBs or P-MBs. We will call these neighboring pixels I-neighbors and P-neighbors, respectively. These pixels are also received because they belong to the same slice as the current MB. The average distortion of P-neighbors is $D_{\mathrm{PR},n}$ by definition. Denoting the average distortion of I-neighbors by $D_{\mathrm{IR},n}^{\mathrm{past}}$, the distortion of the current I-MB can be written as

$$\begin{aligned} D_{\mathrm{IR},n}^{\mathrm{current}} &= E\left\{\left(f_{p,i,e}^i - f_{p,i,d}^i\right)^2\right\} \\ &= E\left\{\left(\sum c_l e^{q_l(i)}\right)^2\right\} \\ &= c_I D_{\mathrm{IR},n}^{\mathrm{past}} + c_P D_{\mathrm{PR},n} \end{aligned} \quad (26)$$

with

$$c_{\mathrm{I}} = E_a\left\{\sum_{l:\mathrm{I-neighbors}} c_l^2 + \rho \sum_{l,k:\mathrm{I-neighbors},l\neq k} c_l c_k\right\} \quad (27)$$

$$c_{\mathrm{P}} = E_a\left\{\sum_{l:\mathrm{P-neighbors}} c_l^2 + \rho \sum_{l,k:\mathrm{P-neighbors},l\neq k} c_l c_k\right\}. \quad (28)$$

In deriving the above result, we have assumed the channel-induced errors in I-neighbors and that in P-neighbors are uncorrelated.[3] Equation (26) shows that the distortion of I-MBs evolves in a recursive manner. This means that intraprediction causes spatial error propagation within the same slice. We will assume that this distortion quickly converges after a few I-MBs so that

$$D_{\mathrm{IR},n}^{\mathrm{current}} = D_{\mathrm{IR},n}^{\mathrm{past}} = D_{\mathrm{IR},n}.$$

Substituting this assumption into (26) and making use of (15) yields

$$D_{\mathrm{IR},n} = \frac{c_P}{1-c_I} D_{\mathrm{PR},n} = \frac{ac_P}{1-c_I} D_{c,n-1}. \quad (29)$$

[3]This assumption may not be strictly true. But our experimental data show that the resulting relation in (30) is quite accurate, with c possibly taking into account the discounted correlation.

Assuming the number of P-neighbors and, hence, c_P are proportional to $1-\beta_n$, and similarly the number of I-neighbors and, hence, c_I are proportional to β_n, the factor $(ac_P)/(1-c_I)$ will have the form of $c(1-\beta_n)/(1-d\beta_n)$, where c and d are two scaling parameters. When β_n is small, this factor can be approximated by $c(1-\beta_n)(1+d\beta_n)$, which equals $c(1-\beta_n^2)$ when d is close to 1. From experimental data, we have found that this approximation with only one parameter works quite well. This leads to[4]

$$D_{\mathrm{IR},n} = c(1-\beta_n^2)D_{c,n-1}. \tag{30}$$

For P-MBs, if it is received, the channel distortion $D_{\mathrm{PR},n}$ is the same as in (15) with generally sub-pel MVs. With intraprediction, received I-MBs do not stop error propagation any more. Therefore, the correlation coefficient between adjacent pixels does not necessarily decrease with the intrarate. Hence, we do not need to apply the more complicated "Model2" for $D_{\mathrm{PR},n}$.

For a lost MB, with motion-compensated temporal concealment, the distortion may stay as (30). Substituting (30), (15), and (20) into (6) yields the same recursion form as (5) with

$$\text{Model3}: \quad \alpha = (1-P)(a(1-\beta_n)+c\beta_n(1-\beta_n^2))+hP. \tag{31}$$

The proceeding analysis assumes unconstrained intraprediction. With constrained intraprediction as in H.264, only intracoded neighboring pixels in the same slice can be used for intraprediction so that $c_P = 0$. Consequently, $c = 0$ and the overall distortion stays the same as in the case without intraprediction, and can be characterized by either Model1 or Model2.

The distortion model using (5) and (31) will be referred to as Model3, which is applicable when the encoder applies nonconstrained intraprediction.

With constrained intraprediction or without intraprediction, the factor α in either Model1 or Model2 is a decreasing function of β. This is because received I-MBs stop error propagation. This is not the case with unconstrained intraprediction. In fact, (31) reveals that α is not a monotonically decreasing function of β. Rather, α reaches its maximum at some intermediate intrarate. This phenomenon is confirmed with our experimental results given in Section IV.

E. With Deblocking Filtering

Deblocking filtering within the encoding loop is an important tool for enhancing visual quality of coded video, and is incorporated in the earlier H.261 standard and the latest H.264 standard. In the H.264 standard, deblocking filtering is applied in the so-called "in-loop" manner, so that the filtered value for a pixel can be used for filtering the following pixels. Let $\check{f}_n^i$ represent the reconstructed value for pixel f_n^i before filtering, and $\hat{f}_n^i$ the reconstructed value after filtering. Mathematically, we can describe the deblocking operation by[5]

$$\hat{f}_n^i = \sum_{l:\text{past}} w_l \hat{f}_n^{r_l(i)} + \sum_{l:\text{future}} w_l \check{f}_n^{r_l(i)}. \tag{32}$$

The index $r_l(i)$ represents the lth neighboring pixel used for the filtering pixel i. The set $\{l:\text{past}\}$ includes neighboring pixels that have been filtered, and the set $\{l:\text{future}\}$ includes those that have not been filtered. Different from the intra and interpredictions discussed earlier, one of the $r_l(i)$ is pixel i itself, and it is in the set $\{l:\text{future}\}$. The filter coefficients w_l are location and content dependent, satisfying $\sum_l w_l = 1$. Typically, stronger filtering is applied along the block boundaries, and over smoother regions of a frame.

In the decoder, if an MB is received, the same filtering is applied to the decoded values $\bar{f}_n^i$. Let $\bar{f}_n^i$ denote the reconstructed value for pixel f_n^i before filtering, and $\tilde{f}_n^i$ after filtering, the filtered value can be expressed as

$$\tilde{f}_n^i = \sum_{l:\text{past}} w_l \tilde{f}_n^{r_l(i)} + \sum_{l:\text{future}} w_l \bar{f}_n^{r_l(i)}. \tag{33}$$

The average distortion for a received MB is

$$\begin{aligned} D_{R,n}^{\text{current}} &= E\left\{\left(\sum_{l:\text{past}} w_l\left(\hat{f}_n^{r_l(i)} - \tilde{f}_n^{r_l(i)}\right)\right.\right. \\ &\qquad \left.\left. + \sum_{l:\text{future}} w_l\left(\check{f}_n^{r_l(i)} - \bar{f}_n^{r_l(i)}\right)\right)^2\right\} \\ &= w_{\text{past}} D_{R,n}^{\text{past}} + w_{\text{future}} \bar{D}_{R,n} \end{aligned}$$

with

$$w_{\text{past}} = E_a\left\{\sum_{l:\text{past}} w_l^2 + \rho \sum_{l,k:\text{past},l\neq k} w_l w_k\right\} \tag{34}$$

$$w_{\text{future}} = E_a\left\{\sum_{l:\text{future}} w_l^2 + \rho \sum_{l,k:\text{past},l\neq k} w_l w_k\right\}. \tag{35}$$

When deriving the above result, we have assumed that the channel-induced errors in the past pixels are uncorrelated with those in the future pixels. Although this assumption may not be strictly true, the model derived based on this assumption is quite accurate, with the final model parameters possibly incorporating this discounted correlation.

Assuming the distortion quickly converges so that $D_{R,n}^{\text{current}} = D_{R,n}^{\text{past}} = D_{R,n}$, we have

$$D_{R,n} = w\bar{D}_{R,n}, \text{ with } w = \frac{w_{\text{future}}}{1-w_{\text{past}}}$$

[4]In an earlier work [8], we have assumed $D_{\mathrm{IR},n} = c(1-\beta_n)D_{c,n-1}$. We have since found the current relation is slightly more accurate

[5]In the INTRA-mode of H.264, the intrapredicted values are based on the encoded values before deblocking filtering. Therefore, deblocking can be considered as a postfiltering operation

where $\bar{D}_{R,n}$ is the distortion for a received MB if no deblocking filtering is applied. Substituting (15) and (30) for $\bar{D}_{R,n}$ for P- and I-MBs, respectively, yields

$$D_{\mathrm{PR},n} = w_P a D_{c,n-1} \tag{36}$$
$$D_{\mathrm{IR},n} = w_I c(1-\beta_n^2) D_{c,n-1} \tag{37}$$

where w_I and w_P are the "w" constants corresponding to I- and P-MBs, respectively. In general, their values differ because different deblocking filters are typically applied for I- and P-MBs.

For a lost MB, deblocking is typically not applied after concealment. Therefore, its distortion stays as (20). If deblocking is applied, it will have an effect of changing the parameter h in (20) to $w_L h$. Hence, the average distortion has the same form as (5) and (31) but with the definition of a and c and possibly h changed to include the multiplicative effect of the deblocking filtering. Depending on the relative magnitude of w_{past} and $w_{\mathrm{future}}, w_I, w_P$ or w_L can be either smaller or greater than 1. Therefore, recursive deblocking filtering can either attenuate or exacerbate error propagation.

F. Summary of the Models for Different Encoder and Decoder Configurations

To summarize, we propose three models. They all follow the same recursion formula in (5), but with different forms for the error propagation factor α. The definitions of α for Model1, Model2, and Model3 are given in (22), (23), and (31), respectively. Model1 and Model2 are applicable when the encoder either does not apply intraprediction or applies constrained intraprediction, with or without deblocking filtering. Model1 has two parameters a and h, and Model2 has three parameters a, b, h. As will be shown by the experimental results, Model2 is more accurate than Model1. Model3 is applicable when the encoder applies nonconstrained intraprediction, with or without deblocking filtering. It has three parameters a, c, h. When the encoder uses integer-pel MVs only and does not apply deblocking filtering, $a = 1, b = 0$, and Model2 reduces to Model1. When the decoder uses integer-pel MVs only for error concealment and does not apply deblocking filtering after concealment $h = 1$. More generally, the model parameters depend on the weighting coefficients used for interprediction, intraprediction, temporal concealment, and deblocking filtering. Because the actual weighting coefficients used are location and content dependent, it is not easy to determine them directly. But they can be obtained by fitting the distortion model to measured channel distortion values under different loss rates P when a video is coded using different intrarates. This will be discussed further in Section IV-B.

G. Model Simplification

The previous analysis assumes that $D_{\mathrm{ECP},n}$ and β_n vary from frame to frame and can be measured accurately. A simplified model results if we assume that these values stay fairly constant and their average values can be predetermined. Let D_{ECP} and β denote the average concealment distortion and intrarate in all of the P-frames, then (5) becomes

$$D_{c,n} = P D_{\mathrm{ECP}} + \alpha(\beta, P) D_{c,n-1} \tag{38}$$

with

$$\mathrm{Model1}: \alpha(\beta, P) = (1-P)(1-\beta)a + hP \tag{39}$$
$$\mathrm{Model2}: \alpha(\beta, P) = (1-P)(1-\beta)(a + b(1-\beta)) + hP \tag{40}$$
$$\mathrm{Model3}: \alpha(\beta, P) = (1-P)((1-\beta)a + c\beta(1-\beta^2)) + hP. \tag{41}$$

Assume frame 0 is coded in the I-mode and has a channel distortion of $D_{c,0}$. Applying (38) recursively yields

$$\begin{aligned} D_{c,n} &= P D_{\mathrm{ECP}}(1 + \alpha + \cdots + \alpha^{n-1}) + \alpha^n D_{c,0} \\ &= P D_{\mathrm{ECP}} \frac{1-\alpha^n}{1-\alpha} + \alpha^n D_{c,0}. \end{aligned} \tag{42}$$

The average distortion over a GoP consisting of 1 I-frame and $N-1$ P-frames is

$$\begin{aligned} D_{c,\mathrm{GOP}} &= \frac{1}{N}\sum_{n=0}^{N-1} D_{c,n} \\ &= \frac{P D_{\mathrm{ECP}}}{1-\alpha}\left(1 - \frac{1-\alpha^N}{N(1-\alpha)}\right) + \frac{1-\alpha^N}{N(1-\alpha)} D_{c,0}. \end{aligned} \tag{43}$$

Because $\alpha < 1$ in general, (42) tells us that $D_{c,n}$ is generally an exponentially increasing function of n and converges to

$$D_c = \lim_{n\to\infty} D_{c,n} = \frac{P}{1-\alpha(\beta, P)} D_{\mathrm{ECP}}. \tag{44}$$

Without intraprediction or with constrained intraprediction, in the special case of integer-pel MVs and no deblocking filtering $(a = h = 1, b = 0), \alpha = 1 - \beta(1-P)$, the converged value is

$$D_c = \frac{P}{\beta(1-P)} D_{\mathrm{ECP}}. \tag{45}$$

In this case, D_c is inversely related to β. For P small, D_c increases with P approximately linearly. Because the values of a, h are close to 1 even with sub-pel MVs and deblocking filtering, the above relation will stay approximately true. The converged form (44) and its special case in (45) are very useful for analytical studies involving channel distortion.

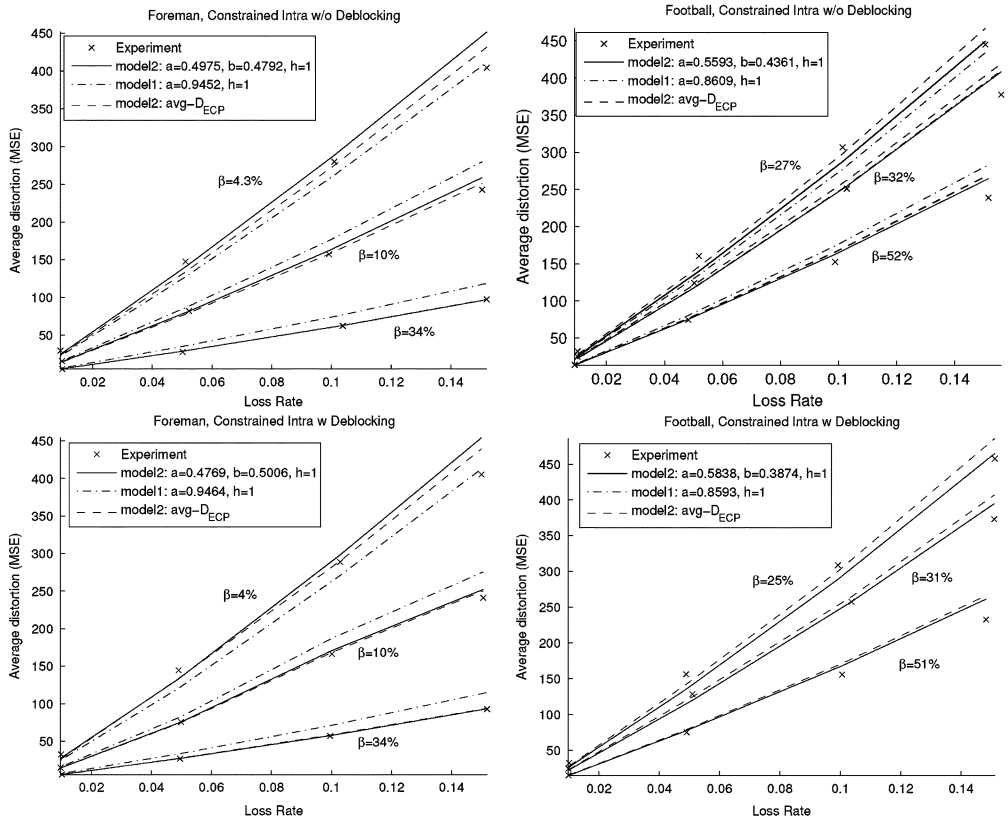


Fig. 1. Average channel distortion over all P-frames versus packet loss rates for different average intrarates. Constrained intraprediction.

IV. Verification of the Model

A. Simulation Setup

The H.264 reference software encoder (JM9.6) with different encoding options in the Baseline profile was used to generate the test sequences. To verify Model1 and Model2, we used "constrained intraprediction" option. To verify Model3, we used the "nonconstrained intraprediction" option. We used one slice per frame so that a lost slice leads to a lost frame. The encoder used up to quarter-pel accuracy for motion estimation, with a motion search range of 16 pels. We also limited the prediction reference frame to only the previous frame. Four encoding configurations were examined: constrained intraprediction without deblocking filtering, constrained intraprediction with deblocking filtering, nonconstrained intraprediction without deblocking filtering, and nonconstrained intraprediction with deblocking filtering.

Two types of error concealment methods were examined: frame copy, which simply replaces a missing frame by the previously reconstructed frame; and motion copy, which copies the motion vector data for the previously reconstructed frame to the current missing frame, and then applies motion-compensated prediction to reconstruct the missing frame. (more details will be given in Section IV-D). Using either concealment method, we determine the concealment distortion for each P-frame by setting that frame alone as lost.

Four common QCIF sequences were examined: two high motion sequences "Football" and "Stefan," one medium motion sequence "Foreman," and one low motion sequence "News." The actual encoding frame rates varied among the different experiments and are described later. For each sequence, only the first 60 frames were coded, where the first frame was coded as an I-frame, while the remaining 59 frames were coded as P-frames using different forced intrarates. A constant $\mathrm{QP} = 28$ is used in all experiments. The P-frame data were subject to random frame losses at rates of 1%, 5%, 10%, and 15%. For a given target loss rate and a forced intrarate, 200 to 500 loss traces are generated. The channel distortion for each frame is determined by averaging the distortion (in terms of MSE) resulting from all loss traces.

B. Model Parameter Estimation

By definition, parameters a and b (14) depend on the distribution of interpolation filters used for sub-pel motion compensation which, in turn, depend on the distribution of the resolutions (integer-pel versus half-pel versus quarter-pel) of the motion vectors used for temporal prediction in INTER mode. Similarly, parameter h (21) depends on the distribution of the resolutions of the motion vectors used for temporal concealment. Finally, parameter c (27)–(29) depends on the distribution of different INTRA predictors, and the percentage of neighboring pels used for intraprediction that are coded in the INTER mode. All of these parameters also depend on the distribution of the deblocking filters applied, and the correlation between the channel-induced errors in adjacent pixels. Deriving the model parameters based on the aforementioned statistics is, however, quite difficult. Instead, we choose to use the least squares fitting technique to derive the model parameters based on the training data.

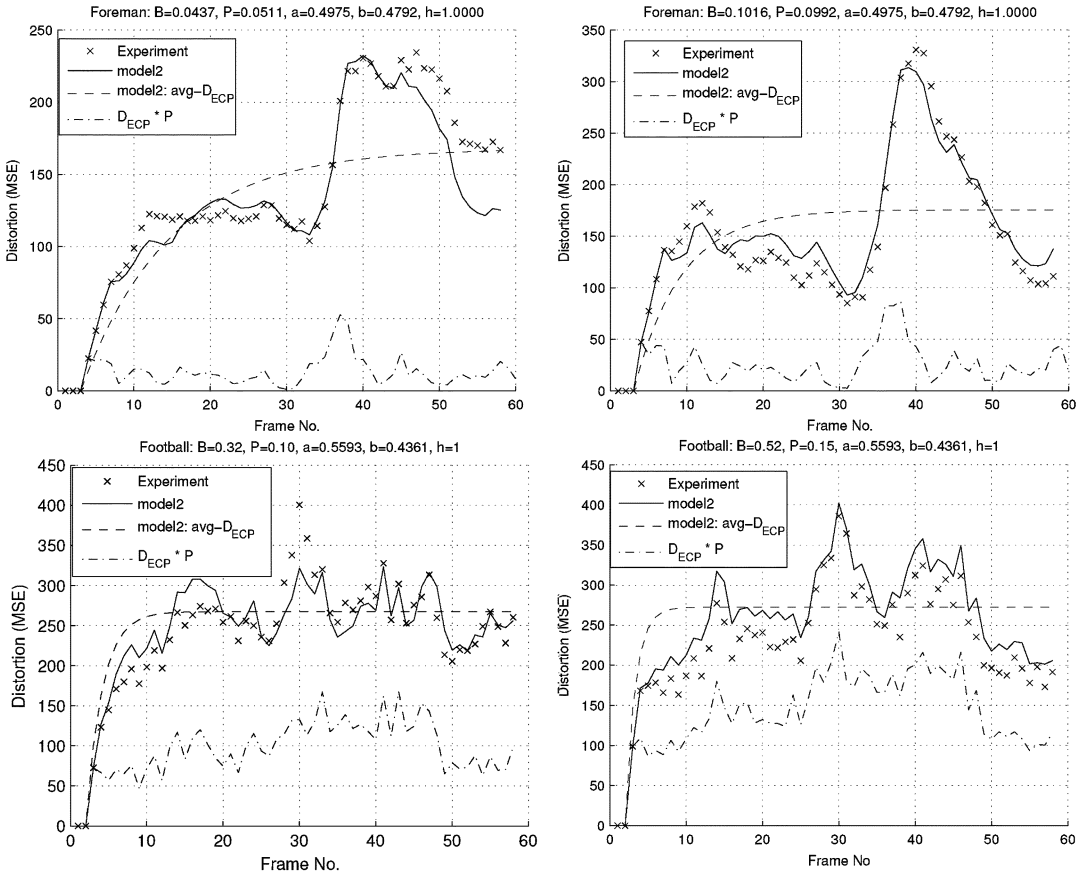


Fig. 2. Distortion versus frames for selected cases. Constrained intraprediction, no deblocking filtering. B indicates the average intrarate, and P the average loss rate.

Specifically, we apply least square fitting to the recursion formula (5) with α defined for different cases, using the data obtained at different loss rates and intrarates. For example, for Model2, given in (5) and (23), the parameters a, b, h are determined by minimizing

$$\sum_{P,\beta_{\rm FI}} \sum_{n} [D_{c,n} - P_n D_{{\rm ECP},n} - ((1-P_n)(1-\beta_n) \times (a + (1-\beta_{n-1})b) + P_n h) D_{c,n-1}]^2 \quad (46)$$

where the sum over $P, \beta_{\rm FI}$ covers all of the tested pairs of loss rate and forced intrarate, and the sum over n considers all P-frames in the decoded sequence obtained with a particular pair of P and $\beta_{\rm FI}$. The terms $D_{c,n}, D_{{\rm ECP},n}, \beta_n, P_n$ denote the measured channel distortion, concealment distortion, intrarate, and loss rate for frame n obtained with this pair of P and $\beta_{\rm FI}$. The parameter estimation for other models are carried out similarly.

C. Results With the Frame-Copy Error Concealment Method

In the first set of experiments, the decoder conceals a lost frame by copying from the previous reconstructed frame. The concealment distortion $D_{{\rm ECP},n}$ is simply the mean squared difference between two encoded frames and can be easily measured. Recall that with frame copy, the parameter $h = 1$ is according to our analysis. So we set $h = 1$ and obtain the other model parameters through least squares fitting as described in Section IV-B. For this experiment, we only tested two sequences "Foreman" and "Football," encoded at 15 and 30 fps. The forced intrarates included 3/99, 9/99, and 33/99. For a given target loss rate and a forced intrarate, 500 simulations were run.

We first examine the results for the case with constrained intraprediction using Model1 and Model2. Fig. 1 shows the average channel-induced distortion over all P-frames versus the packet loss rate. The top two figures are obtained without deblocking filtering, and the bottom two are with deblocking filtering. The model parameters derived in each case are indicated in the figure legend. We see that "Model2" fits the experimental data very well over the large range of loss rates and intrarates examined (less accurate at high loss rates). "Model1," with one less parameter, is less accurate for "Foreman" (but almost as good for "Football"), but still provides a quite good approximation. "Model1" and "Model2" are obtained by using the actual $D_{{\rm ECP},n}$. "Model2-avg-DECP" is computed by using the average concealment distortion over all frames, which is almost as accurate as "Model2." Therefore, the model can estimate the average channel distortion accurately even if we only know the average concealment distortion.

On each figure, different sets of curves correspond to the results obtained using different forced intrarates, and the β values indicated on the figure are the corresponding average intrarates over all P-frames. Notice that with the same forced

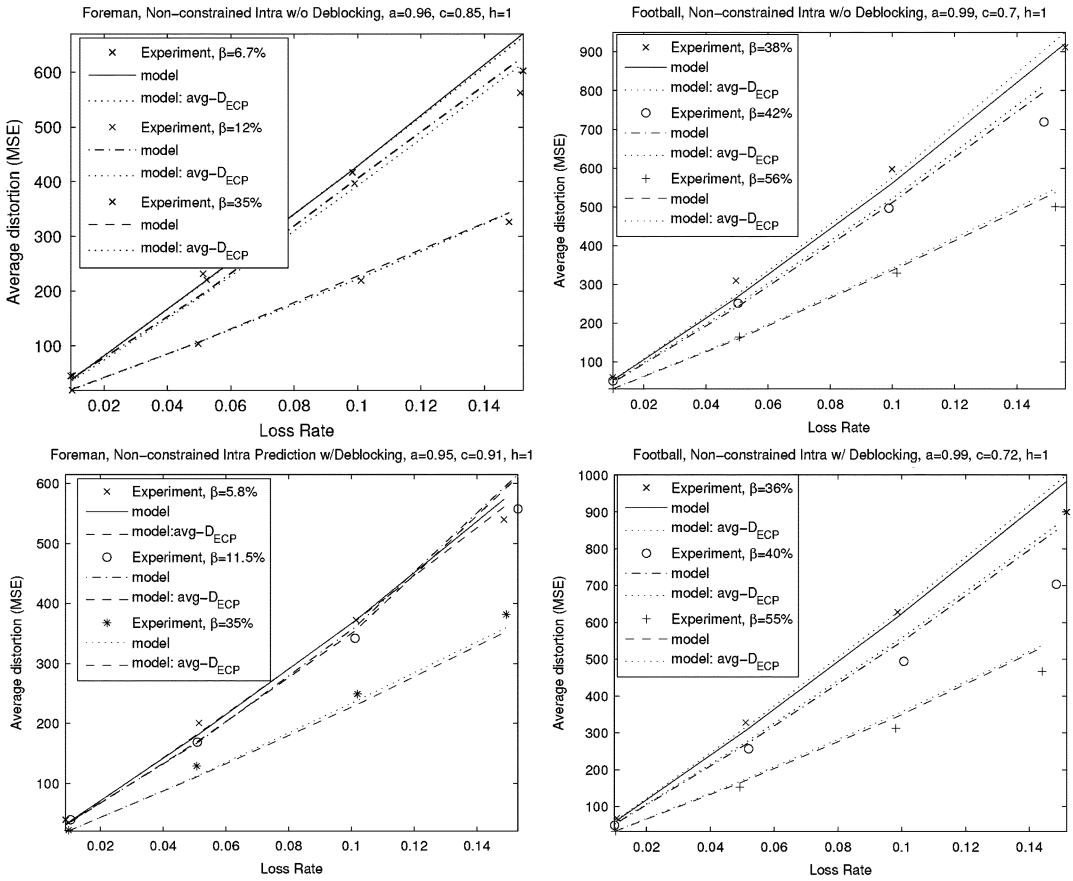


Fig. 3. Average channel distortion over all P-frames versus packet-loss rates for different average intrarates. Nonconstrained intraprediction.

intrarate, the actual intrarate for "Football" is much higher than for "Foreman," because "Football" has a complicated motion and interprediction is not effective for many blocks.

Comparing the results obtained with and without deblocking filtering, we see that the distortion curves and the model parameters are very close in these two cases for the same video sequence. This reveals that the multiplicative factor due to deblocking filtering w_I and w_P are fairly close to 1. As shown in our model derivation, deblocking may either increase or decrease the various model parameters due to the recursive nature of the filtering operation. Therefore, recursive deblocking filtering does Not necessarily further suppress the error propagation.

The previous figure shows that the estimated average distortion over all P-frames in a GOP matches quite well with the measured average distortion. Fig. 2 shows the distortion versus frame number for a few selected pairs of forced intrarate and packet-loss rates. We see that the estimated distortion matches with the actual frame-level distortion quite well.[6] To show the dependency between the channel distortion in a frame with the concealment distortion for that frame, we also show the concealment distortion (scaled down by the packet-loss rate for that frame) versus the frame number curve as well. Recall that if there was no error propagation, the channel distortion for a frame would be equal to the concealment distortion multiplied by the packet-loss rate. The difference between the two curves shows the effect of error propagation due to interprediction. By comparing the two curves, we can also see the influence of the concealment distortion on the total channel distortion.

On the same figure, we also show the estimated distortion based on the average concealment distortion and average intrarate over all P-frames. As expected, the result does not match well with the actual distortion. But as shown in Fig. 1, the average channel distortion over all P-frames based on the model does match well with the measured average distortion. This tell us that if the goal is to estimate the channel distortion in each frame accurately, one needs to estimate or measure the concealment distortion for that frame accurately. But if one is only concerned about the average channel distortion over a video segment, then having an accurate estimate for the average concealment distortion for this segment is sufficient.

Next, we examine results obtained with nonconstrained intraprediction using Model3. Fig. 3 shows that Model3 fits the experimental data quite well, both with the actual concealment distortion in each frame and with their average value. Comparing Figs. 1 and 3, we see that under the same intrarate and packet-loss rate, the channel distortion is much higher (especially for the "Football" sequence) when nonconstrained intraprediction is used. Therefore, for error resilience purposes, constrained intraprediction is much preferred. As with Fig. 1, the distortion curves and the model parameters are very close with and without deblocking filtering. This again confirms that deblocking filtering essentially modifies the model parameters

[6]Only the results with Model2 are shown. The curves corresponding to Model1 follow the same trend, but are slightly less accurate.

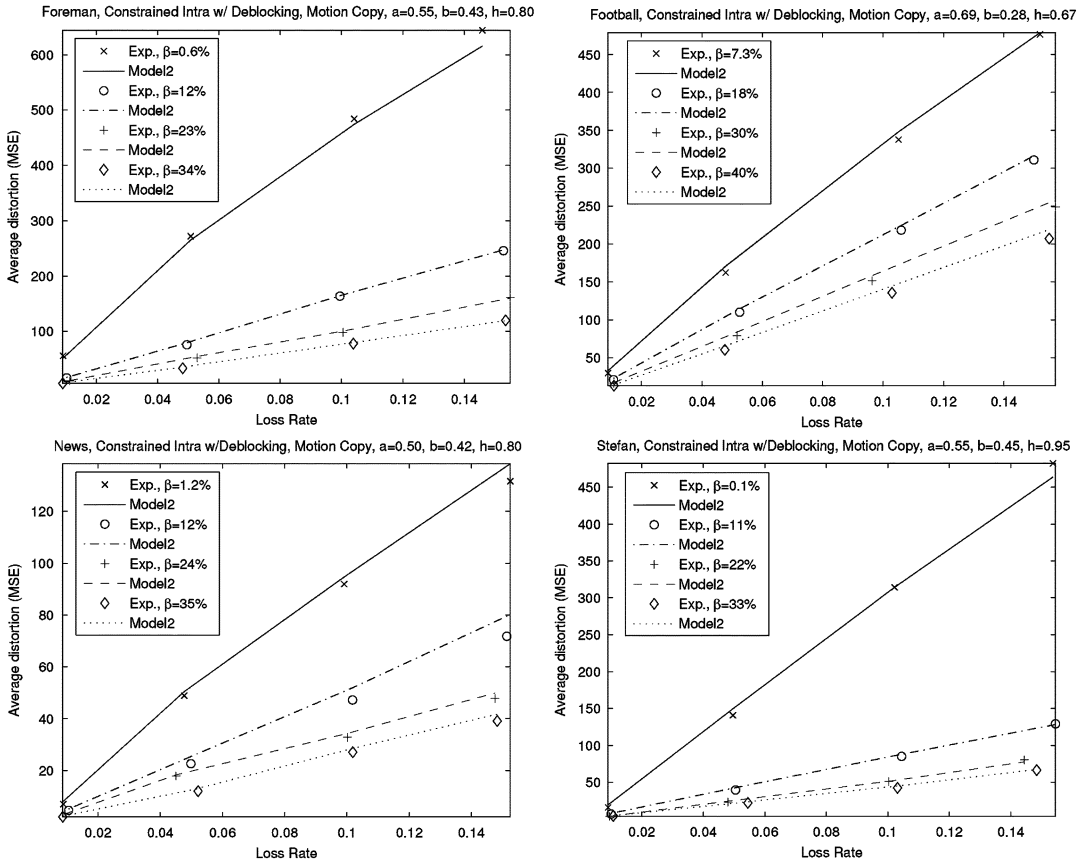


Fig. 4. Average channel distortion over all P-frames versus average loss rates, for different intrarates. Constrained prediction with deblocking filtering. Motion copy error concealment.

slightly. For this set of results, the least squares method did not give the best fit and, for some cases, the derived parameter a exceeds 1. We have manually adjusted the parameters slightly to obtain a better fit with the experimental data.

Note that for "Foreman," the experimental curves corresponding to the first two forced intrarates are very close to each other. This is because with nonconstrained intraprediction, INTRA-coded MBs do not always stop error propagation, unless all of the blocks are coded as INTRA. Therefore, increasing the intrarate does not necessarily decreases channel distortion. In fact, the error propagation factor α, as defined in (31), is a nonmonotonic function of the intrarate, and reaches its maximum at some intermediate intrarate. We will show figures relating the channel distortion to the intrarate for the next set of experiments.

D. Results With the Motion Copy Error Concealment Method

In the second set of experiments, the decoder conceals a lost frame by using the motion information of the previously reconstructed frame. Specifically, each MB in the lost frame is assigned the same mode and motion vectors of its co-located MB in the previous decoded frame. If the co-located MB is coded in an INTRA mode, then the current MB is assigned the SKIP mode, and the motion vector for this MB is estimated from the MVs of its surrounding previously decoded MBs. The decoder then invokes the conventional motion compensation operation to recover each MB [9]. For this experiment, we tested four sequences: "News," "Foreman," and "Stefan," encoded at 10 fps, and "Football," encoded at 30 fps, respectively. The forced intrarates included 0, 11/99, 22/99, and 33/99. For a given target loss rate and a forced intrarate, 200 simulations were run. For each sequence, we derive the model parameters a, b, h or a, c, h through least squares fitting as described in Section IV-B. For this experiment, we examined only the following two cases: constrained intraprediction with deblocking, and nonconstrained intraprediction with deblocking.

Fig. 4 shows the results with constrained intraprediction for all test sequences. To reduce the clutter of the curves, for each intrarate, we only show the experimental data and the modeled result using Model2. The difference between Model2 and Model2:avg-DECP, and that between Model2 and Model1, is very similar to the trend shown in Fig. 1. We see the modeling results are very accurate for all four sequences, at different packet-loss rates and intrarates, as has been shown in the previous experiment (Fig. 1). We see that with motion copy, the parameter h is in the range of 0.65–0.95.

To examine the influence of the intrarate on the channel distortion more closely, Fig. 5 shows the same set of results, but plotting the channel distortion versus the intrarate, for different packet-loss rates. We see that the model was able to accurately track the variation of the channel distortion with the intrarate. The channel distortion basically decreases with the intrarate in the form of $1/\beta$, as suggested by (45). We also see that increasing the forced intrarate from 0 to about 11% greatly reduces the channel distortion, but further increasing the intrarate does not bring in as many significant gains.

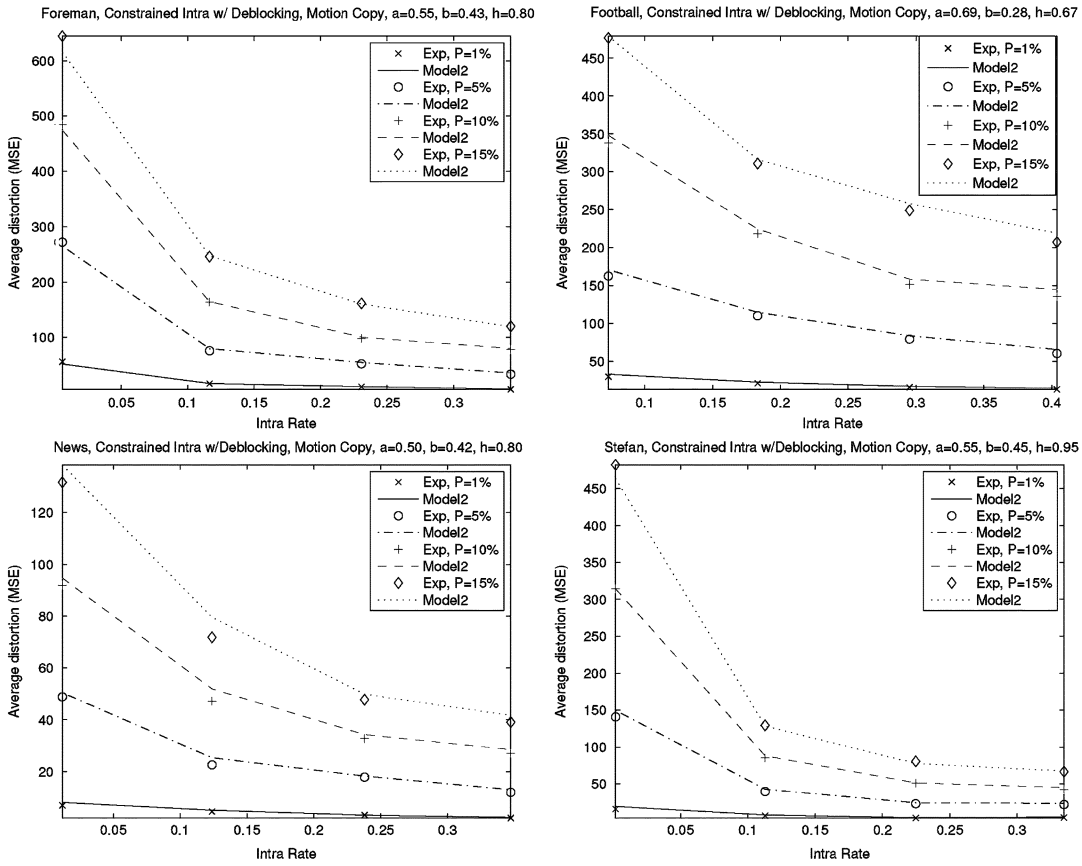


Fig. 5. Average channel distortion over all P-frames versus average intrarates for different loss rates. Constrained intraprediction with deblocking filtering. Motion copy error concealment.

We next show the results with nonconstrained intraprediction. Fig. 6 presents the curves of the distortion versus packet-loss rate. Fig. 7 presents the curves of the distortion versus the intrarate. For this case, the experimental data obtained for some two adjacent intrarates are very close to each other (the middle two intrarates for "Foreman" and "News," the first two intrarates for "Football") and Model3 was not able to reproduce this nonmonotonic relation with the intrarate accurately. But overall, the model is still fairly accurate over the range of the intrarates and packet-loss rates tested.

Recall that with Model3, we model the channel distortion in received INTRA-coded MBs using (29). By assuming the factor $c_P/(1-c_I)$ is proportional to $(1-\beta^2)$, we further approximate (29) with (30). This approximation leads to α defined in (31). Although the so-derived α reflects the general nonmonotonic relation of the distortion with the intrarate, it fails to capture the exact relation. We believe that $c_P/(1-c_I)$ is probably related to β in a more complicated form than assumed, and a more accurate model will require more parameters.

E. Summary of Model Parameters

Table I summarizes the model parameters for different sequences under different encoder and decoder configurations. It can be seen that the parameter a for Model1 (for constrained intraprediction) and that for Model3 (for nonconstrained intraprediction) falls in the range of 0.86–0.99. With Model2 (for constrained intraprediction), the sum of a and $b(1-\beta)$ should be in the same range as a in Model1. For nonconstrained intraprediction, the parameter c is around 0.7 for all sequences except for "Foreman." With motion copy, $h = 0.8$ is a good choice, except for "Stefan," which requires a much higher value. This sequence also requires a higher a value.

In practical applications, it may not be feasible to determine the model parameters based on training data. We had hoped sequences with similar motion content would have similar parameters. However, although "Football" and "Stefan" are most similar in terms of motion content among our test sequences, they have quite different model parameters. Recall that parameters a and h depend on the distribution of the motion vector resolutions, and parameter c depends on the distribution of the intrapredictor. It is possible that these two high-motion sequences have very different distributions in this regard. Ideally, one should categorize video sequences based on these statistics and predetermine the parameters for each category.

V. Channel Distortion Model for Single Frame Loss

The distortion model described so far assumes that packet loss occurs randomly with a loss rate of P. We now extend it to consider the case when only a single frame in a GoP is lost. Assuming one can determine the concealment distortion for any frame if this frame alone is lost, we would like to model the channel distortion in all subsequent frames in the GoP due to the prediction loop in the encoder and decoder. This model allows one to determine the average distortion over the GoP due

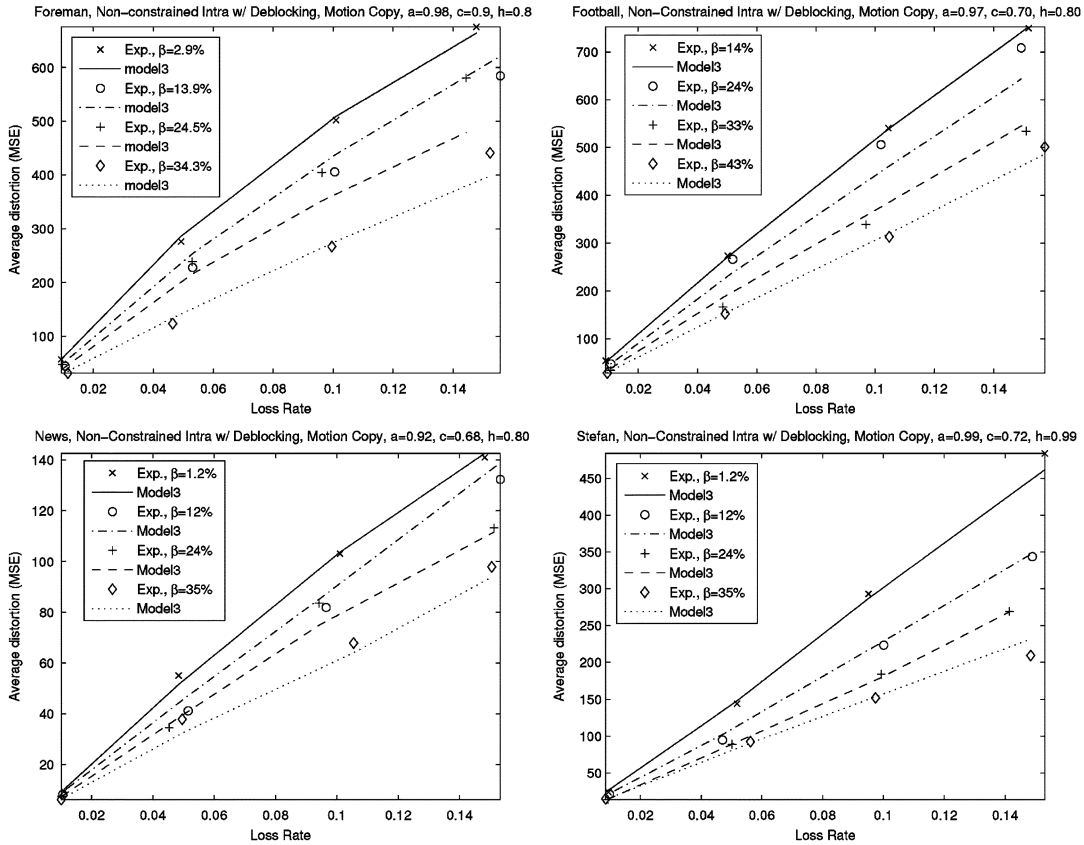


Fig. 6. Average channel distortion over all P-frames versus average loss rates, for different intrarates. Nonconstrained intraprediction with deblocking filtering. Motion error concealment.

to the loss of any one frame in the GoP. Such a measure is useful, for example, when unequal error protection is being considered over different frames. Stronger error protection should be applied to the frames whose loss will lead to higher average distortion.

Assume frame m is lost and concealed with a distortion $D_{\mathrm{ECP},m}$. Due to the use of interprediction, the channel distortion in frame m will be propagated to the following frames, even if they are all received. Denote $D_{c,n;m}$ as the distortion in frame $n(n > m)$ due to the loss of frame m. This distortion can be determined by using (5) with $P = 0$. This yields

$$\begin{aligned} &D_{c,m+1;m} = \bar{\alpha}(\beta_{m+1}) D_{\mathrm{ECP},m} \\ &D_{c,m+2;m} = \bar{\alpha}(\beta_{m+2}) D_{c,m+1;m} \\ &\cdots\cdots \\ &D_{c,n;m} = \bar{\alpha}(\beta_n) D_{c,n-1;m} \\ &= \left(\Pi_{k=m+1}^{n} \bar{\alpha}(\beta_k)\right) D_{\mathrm{ECP},m}, n = m+1, \ldots, N-1 \end{aligned} \tag{47}$$

where $\bar{\alpha}(\beta)$ can be determined using (22) or (23) if constrained intraprediction is used, or (31), if nonconstrained intraprediction is used, with $P = 0$. Specifically, we have

$$\mathrm{Model1}: \quad \bar{\alpha}(\beta_n) = a(1-\beta_n) \tag{48}$$

$$\mathrm{Model2}: \quad \bar{\alpha}(\beta_n) = (1-\beta_n)(a + (1-\beta_{n-1})b) \tag{49}$$

$$\mathrm{Model3}: \quad \bar{\alpha}(\beta_n) = (1-\beta_n)a + \beta_n(1-\beta_n^2)c. \tag{50}$$

When the intrarates of different frames are fairly constant and can be approximated by an average intrarate β, the above recursion is reduced to

$$D_{c,n;m} \approx \bar{\alpha}^{n-m} D_{\mathrm{ECP},m} \tag{51}$$

with

$$\mathrm{Model1}: \quad \bar{\alpha} = a(1-\beta) \tag{52}$$

$$\mathrm{Model2}: \quad \bar{\alpha} = (1-\beta)(a + (1-\beta)b) \tag{53}$$

$$\mathrm{Model3}: \quad \bar{\alpha} = (1-\beta)a + \beta(1-\beta^2)c. \tag{54}$$

Equation (51) tells us that the channel distortion in the subsequent frames decays exponentially. The average distortion over the GoP is

$$D_{c;m;\mathrm{GoP}} = \frac{1}{N}\sum_{n=m}^{N-1} D_{c,n;m} \approx \frac{1-\bar{\alpha}^{N-m}}{N(1-\bar{\alpha})} D_{\mathrm{ECP},m}. \tag{55}$$

Note that $D_{c;m;\mathrm{GoP}}$ reveals the impact of frame m on the average video quality if this frame alone is lost. Equation (55) thus provides a simple way to determine the importance of frame m.

To verify the proposed model for single-frame loss, we conducted another set of experiments, with the test sequences "Foreman" (10 fps) and "Football" (30 fps). We only verified the results when the encoder applies constrained intraprediction. For each sequence, we coded the first 60 frames (one

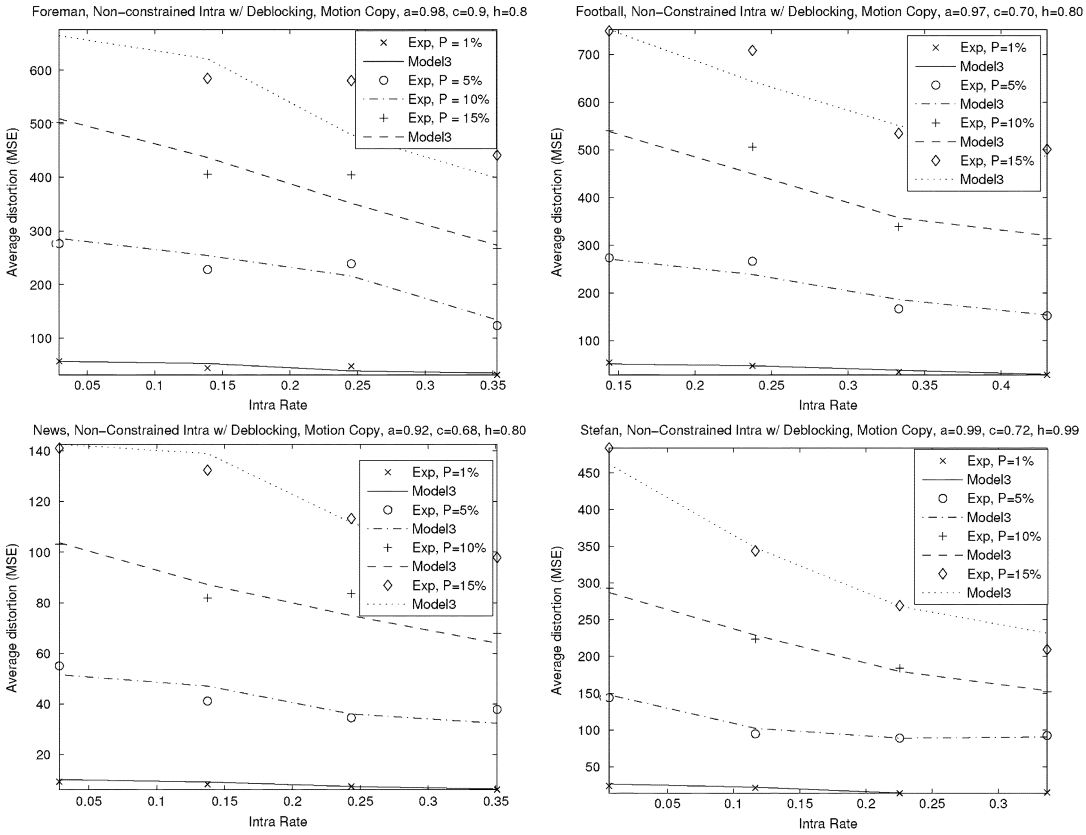


Fig. 7. Average channel distortion over all P-frames versus average intrarates, for different loss rates. Nonconstrained intraprediction with deblocking filtering. Motion copy error concealment.

TABLE I
SUMMARY OF MODEL PARAMETERS FOR DIFFERENT SEQUENCES

Sequence	Encoder Configuration	Concealment	Model	a	b or c	h
Foreman	constrained intra w/o deblocking	frame copy	Model2	0.50	0.48	1.00
			Model1	0.95	0.00	1.00
	constrained intra w/ deblocking	frame copy	Model2	0.48	0.50	1.00
			Model1	0.95	0.00	1.00
	constrained intra w/ deblocking	motion copy	Model2	0.55	0.43	0.80
			Model1	0.96	0.00	0.80
	non-constrained intra w/o deblocking	frame copy	Model3	0.96	0.85	1.00
	non-constrained intra w/ deblocking	frame copy	Model3	0.95	0.91	1.00
	non-constrained intra w/ deblocking	motion copy	Model3	0.98	0.90	0.80
Football	constrained intra w/o deblocking	frame copy	Model2	0.56	0.44	1.00
			Model1	0.86	0.00	1.00
	constrained intra w/ deblocking	frame copy	Model2	0.58	0.39	1.00
			Model1	0.86	0.00	1.00
	constrained intra w/ deblocking	motion copy	Model2	0.69	0.28	0.67
			Model1	0.93	0.00	0.67
	non-constrained intra w/o deblocking	frame copy	Model3	0.99	0.70	1.00
	non-constrained intra w/ deblocking	frame copy	Model3	0.99	0.72	1.00
	non-constrained intra w/ deblocking	motion copy	Model3	0.97	0.70	0.80
News	constrained intra w/ deblocking	motion copy	Model2	0.50	0.42	0.80
			Model1	0.92	0.00	0.80
	non-constrained intra w/ deblocking	motion copy	Model3	0.92	0.68	0.80
Stefan	constrained intra w/ deblocking	motion copy	Model2	0.54	0.45	0.95
			Model1	0.99	0.00	0.95
	non-constrained intra w/ deblocking	motion copy	Model3	0.99	0.72	0.99

GoP) with constrained intraprediction and deblocking filtering turned on. We then set each of the 59 P-frames to be lost and decode the entire GoP with the motion-copy error concealment method. For each chosen index m of the lost frame, we measured the channel distortion of all decoded frames $D_{c,n;m}$. We also recorded the average distortion over the GoP $D_{c;m;\mathrm{GoP}}$. To apply our model, we use the measured $D_{c,n;m}$ data for $m = 1$ to determine a for Model1 or a and b for Model2 using the least squares method. The so-determined parameters yielded a slightly better fit with the experimental data than the parameters derived previously for random packet-loss data.

Fig. 8 compares the experimental data and the modeled results. The top two figures show the distortions of all subsequent frames when the first P-frame is lost. Model1 refers to (47) and (48), and Model2 refers to (47) and (49). For this experiment, we did not force any MBs to be coded in the INTRA mode. For "Foreman," there are very few MBs coded in the INTRA mode so that the actual intrarate is very close to zero in all frames. This caused a problem in finding the correct parameters for Model2 with the least squares method. So we only show the result with Model1. For this sequence, the parameter a is very close to 1, and so is the factor α. This is why the channel distortion decreases quite slowly. For "Football," the parameter a for Model1 (or $a + b$ for Model2) is smaller and we observe the expected exponentially decay behavior. The modeled results with either models follow the measured data quite closely, but consistently underestimated the data. We believe this is because the simple least squares method does not yield the best possible parameters. When we manually increased the parameters slightly, better fitting was obtained.

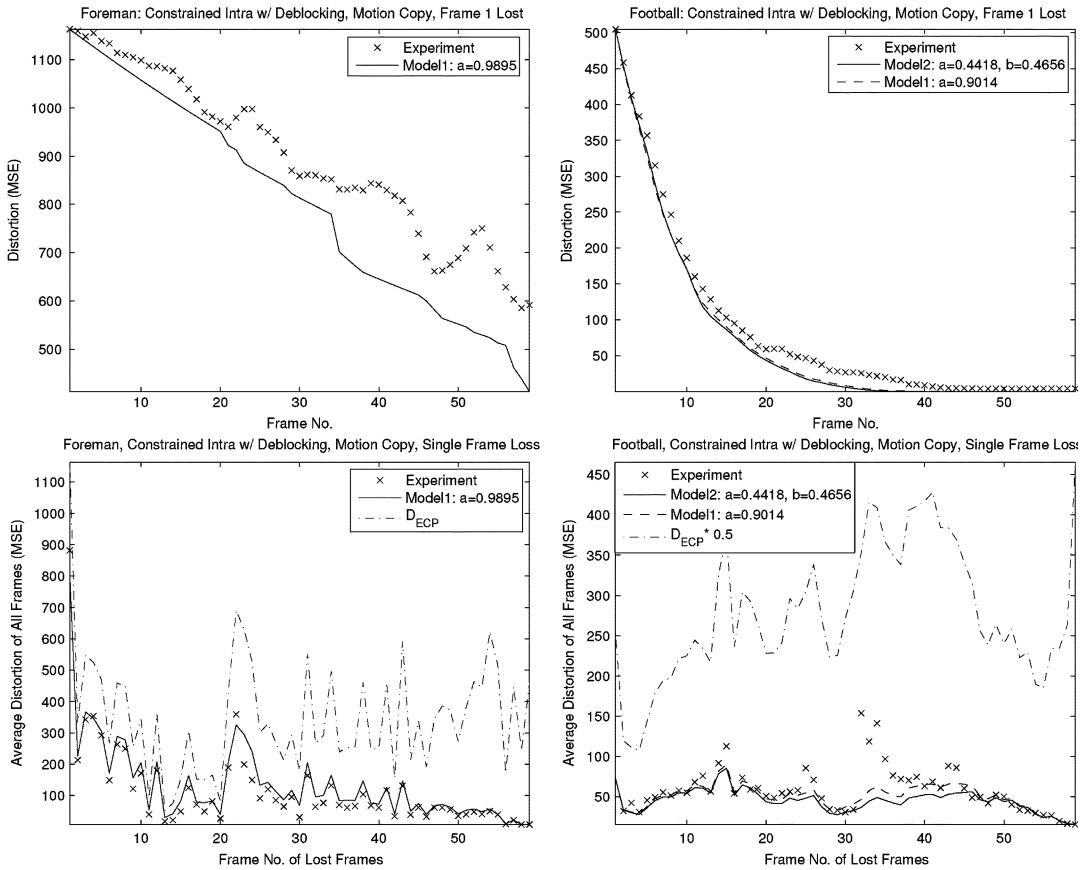


Fig. 8. Channel distortion when a single frame is lost. Constrained intraprediction with deblocking filtering. Motion copy error concealment. Forced intra rate $= 0$.

The bottom two figures in Fig. 8 compare the measured average distortion over 60 frames derived based on our model for all possible loss-frame indices. We see that the model yielded very accurate estimates for most cases. On the same figure, we also plot the $D_{\mathrm{ECP},m}$ data. From (55), the average distortion depends both on the concealment distortion of the lost frame $D_{\mathrm{ECP},m}$ and the location of that frame m. When all frames have the same concealment distortion, a loss that occurs early in a GoP affects more frames in the GOP and, hence, leads to a larger average distortion. In general, $D_{\mathrm{ECP},m}$ is not a constant, so that the average distortion $D_{c;m;\mathrm{GoP}}$ does not always decrease with m. In fact, as shown in the figure, the loss of some later frames with large concealment distortion can lead to higher average distortion.

VI. Conclusion

In this paper, we proposed a mathematical model for the average channel distortion at frame level. The model takes into account the two new features of the H.264 video-coding standard, namely intraprediction and in-loop deblocking filtering. The proposed model represents the current-frame distortion as the sum of the current-frame concealment distortion multiplied by the packet-loss rate, and the previous-frame distortion multiplied by an error propagation factor. This factor depends on the packet-loss rate and the intrarate. Three variations of the model with different error propagation factors were introduced: Model1 and Model2 are for the case when the encoder either does not apply intraprediction or applies only to constrained intraprediction. Model2 is more accurate but requires one extra parameter. Model3 is for the case when the encoder employs nonconstrained intraprediction. We showed that deblocking filtering does not change the functional form of the error propagation factor, rather it merely changes the model parameters slightly. We also discussed how to derive the model parameters based on simulation data. Comparison of the measured channel distortion data from simulations and modeled results showed that Model2 is very accurate over a large range of packet-loss rate (1% to 15%) and intrarate (0 to 33%). Model1 and Model3 are slightly less accurate, but still quite good.

One potential application of the proposed model is in the rate-distortion-optimized configuration of source and channel coders. Given the total channel bandwidth, the proposed channel distortion model, together with appropriate models for source rate and distortion that relate to the source rate and encoder distortion with the intrarate and quantization parameter, enables one to determine the optimal intrarate, quantization parameter, and the channel code rate that will minimize the total distortion.

We also showed how to adapt the proposed distortion model for random packet losses to consider single frame loss. A comparison with simulation data showed that the model for this scenario is also very accurate. In applications where unequal

error protection may be provided to different encoded frames, the most challenging problem is how to rank the frames based on the impact of the loss of a single frame on the overall video quality. The proposed channel distortion model for the single frame loss enables one to easily predict the channel distortion in subsequent received frames once a single frame is lost and, consequently, determine the average distortion over a group of frames when any frame alone in this segment is lost.

To apply either model, one must choose appropriate model parameters. This is the major obstacle for the application of the proposed models. By definition, the model parameters depend on the distribution of motion vector resolutions (integer versus half versus quarter-pel) and the distribution of the intrapredictors. However, categorizing video sequences based on these statistics is not an easy task. We have found that video sequences with high motion can have quite different parameters, probably because the distributions of their motion vector resolutions are quite different. We also found that the least squares method applied to the recursion equation directly does not always yield very good parameters. How to determine the model parameters more accurately and efficiently (without actually running many channel simulations for the underlying sequence) remains an open question.

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, no. 6, 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, vol. 3, pp. 469–472.

[3] 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, no. 6, pp. 952–965, Jun. 2000.

[4] 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, vol. 1, pp. 187–190.

[5] K. Stuhlmueller, N. Faerber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1012–1032, Jun. 2000.

[6] Z. He, H. Cai, and C. W. Chang, "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, no. 6, pp. 511–523, Jun. 2002.

[7] J. Ostermann, "Video coding with H.264/AVC: Tools, performance, and complexity," *IEEE Circuits Syst. Mag.*, vol. 4, no. 1, pp. 7–28, First Quarter 2004.

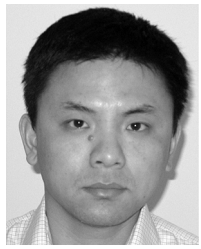
[8] Y. Wang, Z. Wu, J. Boyce, and X. Lu, "Modeling of distortion caused by packets in video transport," in *Proc. IEEE Int. Conf. Multimedia and Expo*, Amsterdam, The Netherlands, Jul. 2005, pp. 1206–1209.

[9] S. Bandyopadhyay, Z. Wu, P. Pandit, and J. Boyce, "Frame loss error concealment for H.264/AVC," in *Proc. ISO/IEC MPEG and ITU-T VCEG, JVT-P072*, Poznan, Poland, Jul. 2005.

Yao Wang (F'04) received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1983 and 1985, respectively, and the Ph.D. degree in electrical and computer engineering from University of California, Santa Barbara, in 1990.

Since 1990, she has been with the Electrical and Computer Engineering faculty of Polytechnic University, Brooklyn, NY. She is the leading author of a textbook *Video Processing and Communications* (Prentice-Hall, 2002).

Dr. Wang received the New York City Mayor's Award for Excellence in Science and Technology in the Young Investigator Category in 2000. She was elected Fellow of the IEEE for contributions to video processing and communications. She was a co-winner of the IEEE Communications Society Leonard G. Abraham Prize Paper Award in the field of communications systems in 2004.



Zhenyu Wu (S'99–M'05) received the B.S. degree in telecommunication engineering and the M.S. degree in signal and information processing engineering from Shanghai University, Shanghai, China, in 1996 and 1998, respectively, and the Ph.D. degree in electrical engineering from the University of Arizona, Tucson, in 2004.

Currently, he is a Technical Staff Member at Thomson Inc. Corporate Research, Princeton, NJ. His research interests include multimedia communications, source and channel coding, and signal processing.

Jill M. Boyce (S'86–M'92) received the B.S. degree in electrical engineering from the University of Kansas, Lawrence, in 1988, and the M.S.E. degree in electrical engineering from Princeton University, Princeton, NJ, in 1990.

Currently, she is Manager of Mobility Systems Research at Thomson Corporate Research, Princeton, NJ. Her research specialties encompass a wide range of video compression and transmission topics, including preprocessing, encoding, decoding, postprocessing, network streaming, and error considerations. Her current primary research focus is enabling video services to mobile devices over wireless networks. She was previously with Lucent Technologies Bell Laboratories, Holmdel, NJ; and AT&T Labs, Holmdel; and Hitachi America, Princeton, NJ. She has actively participated in several international standardization bodies, including ITU-T, MPEG, ATSC, and SMPTE. She was an active contributor to the Joint Video Team (JVT) standardization of ITU-T H.264/MPEG-4 AVC, and had several contributions adopted into the standard. She also chaired JVT breakout groups on B-frames and Weighted Prediction topics and participated in the editing of the H.264/AVC standard. She gave an invited tutorial on H.264/AVC at the International Conference on Consumer Electronics (ICCE) 2005. She has also been an active contributor to the MPEG Scalable Video Coding (SVC) standardization effort currently under development by the JVT. She currently holds 46 granted U.S. patents, with many more pending. She has authored more than 30 conference and journal papers and book chapters, and has made over 30 contributions to international standards bodies.

Ms. Boyce is a member of the IEEE Circuits and Systems Society Multimedia Applications Technical Committee. She was the Video Networking Track Chair and co-organized a tutorial on IPTV for ICCE 2006. Currentky, she is an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.