Computer Communications 32 (2009) 431–438

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Transmission distortion estimation for real-time video delivery over hybrid channels with bit errors and packet erasures

Wenyu Liu, Zhenhua Tang *, Shuang Zhao

Department of Electronics and Information Engineering, Huazhong University of Science and Technology, No. 1037 Luoyu Road, Wuhan 430074, PR China

ARTICLE INFO

Article history:
Received 31 May 2008
Received in revised form 8 November 2008
Accepted 11 November 2008
Available online 24 November 2008

Keywords:
Transmission distortion
Packet erasure
Bit error
Video transmission

ABSTRACT

We study transmission distortion of video delivery over hybrid channels with bit errors and packet erasures. Hybrid channels have been prone to suffering from transmission error in the forms of packet erasures due to congestion at wired networks, as well as bit errors caused by the wireless interference or fading on wireless links. We present a recursion approach to transmission distortion estimation over hybrid channels in successive frames with the prediction mode. This approach is feasible for video transmission applications as it is capable of *online* estimation of transmission distortion on received decoded video. Our extensive simulation results on six well-known video sequences demonstrate that our proposed approach is accurate and robust.

1. Introduction

Providing assess of video services across wireless communication remains a challenging task due to high error rate, time-varying bandwidth, and limited resources on mobile hosts. Accurate distortion estimation of the received decoded video has important applications on efficient optimization and resource allocation at the sender, which is based on rate-distortion (R-D) or joint power and rate-distortion (P-R-D). Resource optimization techniques are crucial for video services across wireless communications [1–5].

The distortion of decoded video at the receiver side, in a video coding and transmission system, is caused by: (1) the source coding distortion by quantization at sender side and (2) transmission distortion at channels. Often these two types of distortion are uncorrelated. The source coding distortion at the sender side can be estimated via empirical or analytic R-D models at the video encoder [6–10]. However, accurate and robust estimation of transmission distortion is still challenging [7] as it involves a variety of factors, such as source characteristics, complicated transmission conditions, and error propagation, just to name a few.

Recent years, many works on this field of transmission distortion (or end-to-end distortion) estimation have been proposed. Broadly, they can be classified into two categories. The first category focuses on statistical approaches. Zhang et al. [11] proposed a recursive optimal per-pixel estimation (ROPE) method, in which the encoder recursively calculates decoded video distortion for each pixel due to quantization and transmission errors. Yang and Rose [12] extended ROPE method to the half-pixel motion compensation applications.

The second category focuses on analytic models. An analytic model, proposed in [6], used a linear filter to describe the channel error propagation behavior taking into account the intra refresh rate and the spatial filtering. In [7], a frame-level recursion channel estimation model was presented, which considered channel distortion propagated from frame to frame and the differences between adjacent frames of the original sequence. Wang et al. [13] developed a channel distortion model through recursion estimation at the frame level integrating intraprediction and deblocking in H.264/AVC video coding standard. A transmission distortion model by using control system approach, proposed in [14], is based on the fading behavior of the impulse transmission errors.

However, all the above-mentioned approaches [7,11–14] focus on the effect of packet loss to the channel distortion (or end-to-end distortion). Specially, in wireless transmission, some of works [7,13,14] assumed that if a bit error is detected in the video packet, the whole packet will be discarded. Thus, it can hardly integrate bit-level error-resilient techniques [15] or unequal error protection (UEP) [16] into transmission distortion modeling and estimation in these works. It is desirable to establish a channel distortion (or end-to-end distortion) estimation method based on the video coding bitsteam.

Sabir et al. [17] modeled joint source and channel distortion for JPEG image in terms of entropy DPCM coding and bit errors in the coding stream. It is extended in [18] to MPEG-4 video transmission, whereas the parameters of the model must be determined by running many simulations for the video database known beforehand.

* Corresponding author. Tel.: +86 27 87792462 405.
E-mail addresses: liuwy@mail.hust.edu.cn (W. Liu), tangedward@126.com (Z. Tang), zhaoshuang208@sina.com (S. Zhao).

0140-3664/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2008.11.015

Thus, such a model may not be suitable for real-time video communication systems or the applications that the video contents are unknown beforehand, such as videoconference.

Many video applications in wireless communication environments such as wireless Internet need video transmission over both the wired networks and the wireless links. Consequently, channel transmission errors consist of packet loss due to congestion in wired networks and packet loss or bit errors due to interference and fading in wireless links [19–23]. Unfortunately, few studies on channel (or end-to-end) distortion estimation for video transmission have taken into account both packet erasures due to congestion and bit errors caused by interference and fading.

In this paper, we present a transmission distortion estimation algorithm, which explicitly considers transmission errors on the hybrid channels, including packet erasures in wired networks and bit errors in wireless links. This estimation method is applicable for video transmission system combined wired and wireless networks. One work close to ours is [18]. They also consider bit-level error resilience and the exact effect of bit errors in the coded bitstream. However, we emphasize the differences from their work. First, we employ a recursion method in successive frames with prediction mode and thus no extra parameters should be determined offline. Second, our proposed method can recursively estimate the transmission distortion of received decoded video for each macroblock (MB) in terms of video packetization information and channel conditions information.

The remainder of this paper is organized as follows. Section 2 introduces the framework of our system. In Section 3, we describe our proposed transmission distortion estimation approach. We present simulation results in Section 4. Section 5 concludes this paper.

2. Framework of system

In this section, we describe the framework of a general video coding and communication system, including the video codec, packetization scheme, hybrid channels, packet loss detection, and error concealment scheme, shown in Fig. 1.

2.1. Video codec

In the MPEG-4 [24] simple profile (SP), there are two basic types of video object plane (VOP): intra-coded VOP and inter-coded VOP, denoted by I-VOP and P-VOP, respectively. For an I-VOP, every MB is coded using the intraframe mode. While MBs are coded with either the interframe prediction mode or intraframe mode in a P-VOP. Here we will refer to MBs coded with the intraframe mode as INTRA-MBs, while MBs coded using the interframe prediction mode as INTER-MBs. To be convenient, we neglect DC/AC prediction in the INTRA-MBs.

2.2. Packetization

We implement two error resilience tools suggested by the MPEG-4 standard [24]: resynchronization and data partitioning, as shown in Fig. 2. We assumed that each packet is composed of integer number of MBs and all of the MBs in a packet belong to the same VOP. It is noted that the packet of an I-VOP shown in Fig. 2(a) does not use data partitioning.

2.3. Error concealment

When a video packet is lost in wired networks, all the MBs of the packet are corrupted. In this scenario, we assume that all the MBs' coefficients are set zero for an I-VOP packet, whereas MBs in the P-VOP packet are replaced with the corresponding MBs at the same location from the previous decoded frame. We further assume that the decoder can detect error coefficients within a bitstream scope. When a bit error is detected in the packet header, the whole packet is damaged. The concealment scheme is similar to the case that packets are lost in wired networks. For an I-VOP packet, if the decoder detects the first bit error in a certain MB, this MB and the following MBs in the packet are decoded as zeros. For a P-VOP packet, as the first bit error is found in the MV of a MB, this MV and the sequential MVs are discarded. To conceal the errors, the decoder will copy the corresponding MBs at the same location from the previous decoded frame. While errors occur only in the motion marker or the texture information, all the texture information of the packet are set zero. In addition, the decoder still implements motion compensation with the correct MV for each MB in the packet.

2.4. Packet loss detection

At the receiver, the system may first check the loss of packets on account of transmission in wired networks. We confirmed that it can be implemented by detecting the sequential numbers of the packets flagged on other layers in practical communication systems.

2.5. Hybrid channels

In this paper, hybrid channels composed of wired networks and wireless links are considered. In practical communication systems, bit errors or packet erasures tend to be burst. But interleaving on the bit-level [25] or the packet-level [26] could make the burst errors spread random even though the delay is trivial in real-time video applications. For simplicity, we assume that bit errors or packet erasures occur randomly.

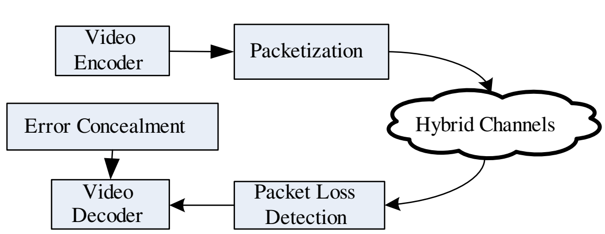


Fig. 1. Framework of video coding and communication system.

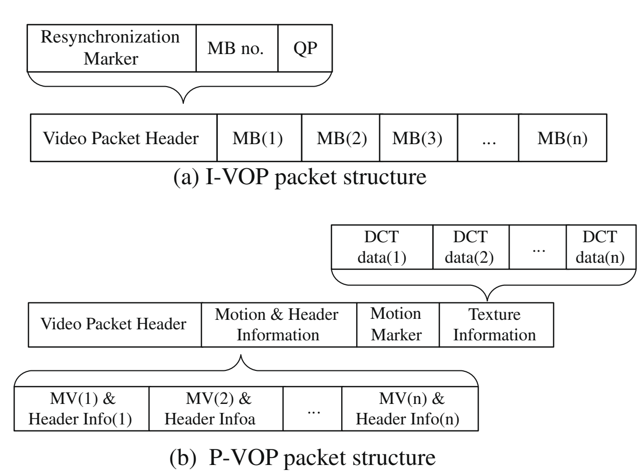


Fig. 2. Bitstream structures for a video packet.

3. Transmission distortion analysis and estimation

In this section, we turn to the transmission distortion of delivering packet video over hybrid channels. We measure the transmission distortion in terms of mean square error (MSE).

Let $\widehat{V}_{f,m,i}$ and $\widetilde{V}_{f,m,i}$ denote the reconstructed signal value at the encoder and at the decoder, respectively. They represent the values of the component i belonging to m^{th} MB in the f^{th} frame. Note that the component is either the luminance or chrominance component. Then the expected transmission distortion for component i is $E_t[(\widehat{V}_{f,m,i} - \widetilde{V}_{f,m,i})^2]$, where $E_t[\cdot]$ represents the expectation of all transmission implementations. We use $E_m\{\cdot\}$ to denote the average distortion over all components in a MB. Accordingly, the average transmission distortion for the m^{th} MB in the f^{th} frame is

$$D_{f,m} = E_m\{E_t[(\widehat{V}_{f,m,i} - \widetilde{V}_{f,m,i})^2]\}. \tag{1}$$

3.1. Distortion estimation for INTRA-MBs

The transmission distortion for INTRA-MBs over the hybrid channels is analyzed in this subsection. Without loss of generality, let the current MB index in the video packet be k.

For the current INTRA-MB, the transmission distortion only results from its loss. As discussed in Section 2.3, the loss of current INTRA-MB possibly caused by packet loss in wired networks, corruption of the packet header, damage of previous MBs or the current one due to bit errors. Note that they will not happen simultaneously. For example, as a video packet has been lost in wired networks, it cannot reach the receiver and the latter cases are impossible. Next we derive the probabilities for them.

It is obvious that erasure of the packet on the wired portion results in loss of the current MB, so the loss probability for the current MB equals the packet erasure rate, denoted by ρ_{plr}. On the other hand, the receiver can get the current packet even though it suffers bit errors in wireless links. Generally, the decoder hardly finds a bit error at the actual position in the variable length coding stream exactly. However, the probability that the decoder can detect a bit error at the exact DCT coefficient is up to 80.4% without considering the cases at later coefficients nearby [17]. Accordingly, we emphasize that the assumption is reasonable that the decoder can find error coefficients within the scope of an average size of MBs in a packet.

Let L_h be the header size of a video packet, then the probability that at least one bit error is detected in the packet header is $P_h = (1 - \rho_{plr})[1 - (1 - \rho_{ber})^{L_h}]$, where ρ_{ber} refers to the bit error rate in the wireless portion. We consider the probability for another case that the packet header is error free while bit errors occur in the segment of MBs. Let L_p and N_{mb} represent the total coding length and the total number of MBs in a packet, respectively. Then the average coding length of MBs can be given by $L_{mb} = (L_p - L_h)/N_{mb}$. If the probability for at least one bit error existing in a MB is $P_{mb} = 1 - (1 - \rho_{ber})^{L_{mb}}$, then the probability for the first bit error occurring in the n^{th} MB can be derived as $P_{mb}(n) = (1 - P_{mb})^{n-1}P_{mb}$. As mentioned in Section 2, bit errors in previous MBs also make the current one corrupted. As a result, the probability for bit errors causing loss of the current MB is $P_{mbloss}(k) = (1 - \rho_{plr})(1 - P_h)\sum_{n=1}^{k} P_{mb}(n)$.

As the current INTRA-MB is corrupted, it is set zero by the decoder. Consequently, the transmission distortion is

$$D_{f,m}(I) = E_m\{E_t[(\widehat{V}_{f,m,i} - 0)^2]\} = (\rho_{plr} + P_h + P_{mbloss}(k))M_{f,m}, \tag{2}$$

where

$$M_{f,m} = E_m[(\widehat{V}_{f,m,i})^2] \tag{3}$$

and here $E_m[(\widehat{V}_{f,m,i})^2]$ denotes the mean square value of an INTRA-MB. From Eqs. (2) and (3), we can see that in order to estimate the transmission distortion of an INTRA-MB over hybrid channels, the packetization information, the channel conditions and the reconsturcted MB at the encoder should be obtained. For the sender, the channel conditions such as bit error rate and packet loss rate can be acquired through a feedback channel [7,21], and the reconstructed signal value of the INTRA-MB and the packetization information are available at the encoder.

3.2. Distortion estimation for INTER-MBs

In this section, we present the transmission distortion estimation approach for INTER-MBs. As a MB is encoded with interframe prediction mode, the MB is predicted by a corresponding MB in a previous frame. Thus transmission errors on the prediction MB in the previous decoded frame may propagate to the current INTER-MB. In addition, loss of an INTER-MB's information such as MV or texture would also cause distortion. Hence, the transmission distortion for the current INTER-MB mainly consists of two parts: concealment distortion introduced by information loss and distortion due to error propagation from the previous frame. Next we discuss the transmission distortion caused by different reasons.

Similar to INTRA-MBs, loss of the current packet in the wired portion or damage of the packet header due to bit errors could cause the whole INTER-MB lost. And the probabilities of this two cases have been described in the previous subsection. We consider another case that the packet header is error-free whereas some bit errors are detected in the segment of MVs. Let L_{mv} represent the average size of MVs in a video packet, then $L_{mv} = L_{mvt}/N_{mb}$, where L_{mvt} denotes the total size of all the MVs in the packet. The probability that at least one bit error occurs in the MV segment of a MB is $P_{mv} = 1 - (1 - \rho_{ber})^{L_{mv}}$, thus the probability for the first bit error occurring in the MV segment of the nth MB in a packet is $P_{mv}(n) = (1 - P_{mv})^{n-1}P_{mv}$. As the first bit error is found in the MVs of previous MBs or the MV of the current one, the current INTER-MB will be corrupted, and the probability is $P_{mvloss}(k) = (1 - \rho_{plr})(1 - P_h)\sum_{n=1}^{k} P_{mv}(n)$.

If the current INTER-MB is damaged, the decoder will copy the MB at the same location from the previous decoded frame to replace it. Note that the MB at the same position in the previous decoded frame may have distortion due to transmission errors. By doing so, this concealment operation would introduce two types of distortion: concealment distortion and distortion from the copied MB. We define $S_{f,m}$ as the concealment distortion, while $D_{f-1,m}$ as the transmission distortion of the MB at the same position in the previous decoded frame. These two types of distortion are obviously uncorrelated. As a result, the transmission distortion caused by loss of the whole current INTER-MB is

$$D_{mbloss}(P) = (\rho_{plr} + P_h + P_{mvloss}(k))(S_{f,m} + D_{f-1,m}). \tag{4}$$

By replacing $\widehat{V}_{f,m,i}$ with $\widehat{V}_{f-1,m,i}$, the concealment distortion can be written as $S_{f,m} = E_m[(\widehat{V}_{f,m,i} - \widehat{V}_{f-1,m,i})^2]$. Here, $E_m[(\widehat{V}_{f,m,i} - \widehat{V}_{f-1,m,i})^2]$ refers to the mean squared error of the MBs at the same location between two successively encoded frame. On the other hand, $D_{f-1,m}$ can be expressed by $E_m\{E_t[(\widehat{V}_{f-1,m,i} - \widetilde{V}_{f-1,m,i})^2]\}$.

In terms of data partitioning, bits of MVs and texture are separated by the motion marker. A possible instance that the segment of MVs is received successfully while bit errors occur in the portion of the motion marker or texture information would make the whole texture information of the packet lost. Let L_{re} and L_{mbm} are the size of texture information and the motion marker, respectively, then the probability for this case is $P_{re} = (1 - \rho_{plr})(1 - P_h)(1 - \rho_{ber})^{L_{mvt}}[1 - (1 - \rho_{ber})^{L_{mbm}+L_{re}}]$. Moreover, as the first bit error occurs in the MVs behind that of the current MB, the following segments including the texture information of the current MB will be also lost. And its occurrence probability is

$$P_{mvnoerr} = (1 - \rho_{plr})(1 - P_h) \sum_{n=k+1}^{N_{mb}} P_{mv}(n).$$

In the above two scenarios, the damaged texture information of the current MB will be set zero, while the correctly received MV is still used to motion compensation. Hence, this will also introduce two kinds of distortion: distortion caused by corruption of the texture information and transmission distortion from the prediction MB in the previous decoded frame. Let $\xi_{f,m}$ and $D_{f-1,m+\Delta d}$ represent the above two types of distortion, respectively, where Δd denotes the motion offset between the current MB and the prediction MB. Then the transmission distortion for the current INTER-MB due to loss of texture information is given by

$$D_{reloss}(P) = (P_{re} + P_{mvnoerr})(\xi_{f,m} + D_{f-1,m+\Delta d}). \tag{5}$$

Let $\hat{e}_{f,m,i}$ represent the residual error component for the current INTER-MB, and then we have $\xi_{f,m} = E_m[(\hat{e}_{f,m,i})^2]$. On the other hand, the transmission distortion of the prediction MB $D_{f-1,m+\Delta d}$ can be obtained by $\sum \partial_j D_{f-1,m+j}$, where ∂_j refers to the proportion that the related MBs in the previous frame which predict the current MB. It is noted that the proportion ∂_j are calculated at half-pixel precision.

Though the MV and texture information of the current MB are received correctly, it is possible that the transmission distortion of the prediction MBs from the previous decoded frame would spread to the current MB. The fact that the current INTER-MB does not suffer from any errors implies the current packet is transmitted error free. Thus, the probability for this case is $P_{errfree} = (1 - \rho_{plr})(1 - \rho_{ber})^{L_p}$, and the transmission distortion for the current MB in this scenario is

$$D_{errfree}(P) = P_{errfree} D_{f-1,m+\Delta d}. \tag{6}$$

From the above analysis, it is easy to find that all the events would contribute to the transmission distortion for the current INTER-MB and the probabilities of them are not interlaced. Finally, the total transmission distortion for the current INTER-MB is the sum of distortion induced by all of the above cases. This can be expressed by adding Eqs. (4)–(6):

$$\begin{aligned} D_{f,m}(P) &= D_{mbloss}(P) + D_{reloss}(P) + D_{errfree}(P) \\ &= (\rho_{plr} + P_h + P_{mvloss}(k))(D_{f-1,m} + S_{f,m}) \\ &\quad + (P_{re} + P_{mvnoerr} + P_{errfree})D_{f-1,m+\Delta d} + (P_{re} + P_{mvnoerr})\xi_{f,m}. \end{aligned} \tag{7}$$

From Eq. (7), we can see that during the process of estimating the transmission distortion for an INTER-MB, the estimation algorithm needs the MSE of the MBs at the same location of the reconstructed frames the texture information of the INTER-MB and the packetization information for the bitstream, which are available at the encoder. Moreover, the estimation algorithm utilizes the estimated transmission distortion of MBs in the previous decoded frame to calculate the distortion propagated to the current INTER-MB. This may lead to a recursion process. We emphasized that, unlike the work in [18], our proposed method does not requires performing channel processes over video sequences to determine the model parameters.

3.3. Transmission distortion estimation for VOPs

We assume the total number of MBs in each VOP is M, then the transmission distortion for I-VOP is the mean value of all the intra-coded MBs' distortion. Thereby, this value can be expressed by

$$D_f(I) = \frac{1}{M} \sum_{m=1}^{M} D_{f,m}(I). \tag{8}$$

In the same way, the transmission distortion for P-VOP is the average distortion of all the INTER-MBs and INTRA-MBs, that is

$$D_f(P) = \frac{1}{M} \left(\sum_{mp=1}^{M1} D_{f,mp}(P) + \sum_{mi=1}^{M2} D_{f,mi}(I) \right), \tag{9}$$

where M is the sum of $M1$ and $M2$.

3.4. Computational complexity analysis

The computational complexity of estimating the transmission distortion of an INTRA-MB mainly concentrates on the calculation of probabilities and mean square values of the reconstructed signal value at the encoder. From the above discussed, we can see that calculating the probabilities of possible cases only deals with several arithmetical operations, and the operations for computing mean square values of the reconstructed INTRA-MB at the encoder are obviously less than the ones of implementing the whole intra coding during video encoding.

For estimating the transmission distortion of an INTER-MB, the major complexity focus on the following aspects: computation of probabilities, the MSE between the current MB and the MB at the same location in the previous frame, mean square values of residual errors, and the transmission distortion of the prediction MB in the previous frame. In fact, the probabilities calculation for INTER-MB also only involves a few arithmetical operations although its complexity is more than the one of INTRA-MB. On the other hand, the operations for calculating the MSE between two MBs and the mean square value of residual errors are hardly ignored compared with the ones of implementing motion prediction for the corresponding INTER-MB at the encoder. In addition, the complexity of computing the transmission distortion of the prediction MB in the previous frame is also less than the one of carrying out motion compensation processing for the INTER-MB since computing the proportion that the related MBs in the previous frame to predict the current INTER-MB by using corresponding MV only involves several arithmetical operations.

It can be seen that the computational complexity of transmission distortion estimation for either INTRA-MB or INTER-MB are both much lower than the one that implements the corresponding encoding for them at the encoder. Hence, the proposed approach can reach the requirement of real-time video communication.

4. Experiment results

In this section, we present the performance results of our proposed approach. We conduct all the experiments using the MPEG-4 [24] verified model (VM).

4.1. Simulation results for given channel conditions

We encoded six QCIF (176×144) video sequences at 15 frames per second, with the first VOP as I-VOP followed by 29 P-VOPs in each group of picture (GOP). These videos have various motion characteristics from high motion to low motion. The total number of the frames is 150 for each sequence. The target source rate and the packet size are set 96 kbps and 1000 bits, respectively. And the size of packet header and the motion marker are assumed to be 32 and 17 bits, respectively. Tables 1 and 2 list different channel conditions for each estimation. And the average packet loss rates of wired networks varies from 4.2% to 21.2%, while the bit error rates of wireless links range from 1.2×10^{-2} to 1.4×10^{-6}. For each video sequence, 50 different random loss patterns are simulated

Table 1
Hybrid channels' conditions for each video and the relative estimation error (REE).

Sequence	Packet loss rate (%)	Bit error rate	REE (%)
News	4.2	1.2×10^{-2}	3.52
Salesman	8.2	1.2×10^{-3}	1.46
Coastguard	10.7	1.2×10^{-4}	1.31
Foreman	12.3	1.0×10^{-5}	3.62
Stefan	14.9	1.3×10^{-6}	4.24
Football	21.2	1.4×10^{-6}	8.96

for given the same target packet loss rate and the bit error rate. Then the average value of 50 runs is considered as the actual distortion value. We use relative estimation error (REE) [14] to evaluate the estimation performance. The REE for each video can be calculated using Eq. (10), where $\widetilde{D}_n$ and $\widehat{D}_n$ are the estimated distortion and the actual distortion, respectively, N is the total number of the encoded frames. Note that $\widetilde{D}_n$ and $\widehat{D}_n$ are peak signal noise ratio (PSNR) values in this paper. The simulation results listed in Tables 1 and 2 show that the RRE is less than 5% in most cases. This indicates that the proposed transmission distortion estimation method is accurate for different channel conditions and video sequences.

$$e = \frac{\sum_{n=1}^{N} |\widehat{D}_n - \widetilde{D}_n|}{\sum_{n=1}^{N} \widehat{D}_n} \times 100\%. \quad (10)$$

In Fig. 3, we plot the average estimated transmission distortion values and the actual ones with different target source coding rates for *mobile* and *akiyo* videos. The encoder settings are the same to the above six videos except for the target source coding rate. Fig. 3(a) shows the results for *mobile* video with the target packet loss rate 5% and bit error rate 10^{-5}, while the results for *akiyo* video with target packet loss rate 10% and bit error rate 10^{-4} is shown in Fig. 3(b). The actual distortion is also the average value for 50 different simulations. It can be seen that the estimated distortion values for different target source coding rates are very accurate within 1 dB compared with the actual distortion values. In fact, more other sequences with different source characteristics are tested and these results are not shown here due to lack of space. The average absolute estimation errors for different channel conditions, videos and source coding rates are not more than 1 dB by employing the proposed transmission distortion estimation method. In addition, we assume that either packet erasure or bit error distribute randomly in the coding bitstream. Given the same target error rate, no matter what target source rates the video sequences are encoded in, they are probable to suffer the equivalent transmission distortion effects roughly. But because of the randomness of channel conditions, our average transmission distortion values obtained by limited channel implementations would be different from the real actual ones. Therefore, the actual values of the reconstructed video quality in different source rates may fluctuate.

In summary, the estimation precision for our presented method is at the same level compared to the work in [18] which estimates distortions within 1 dB in terms of PSNR values in an average manner. However, we consider more error patterns into transmission distortion estimation. First, we take packet loss due to congestion in wired networks into account. Second, the effect that bit errors

Table 2
Relative estimation error (REE) for video *salesman* over hybrid channels.

Sequence	Salesman				
Bit error rate	1.2×10^{-2}	1.2×10^{-3}	1.2×10^{-4}	1.3×10^{-5}	1.2×10^{-6}
Packet loss rate	4.2%	10.2%	12.1%	14.8%	20.0%
REE	3.60%	0.91%	4.56%	1.21%	0.58%

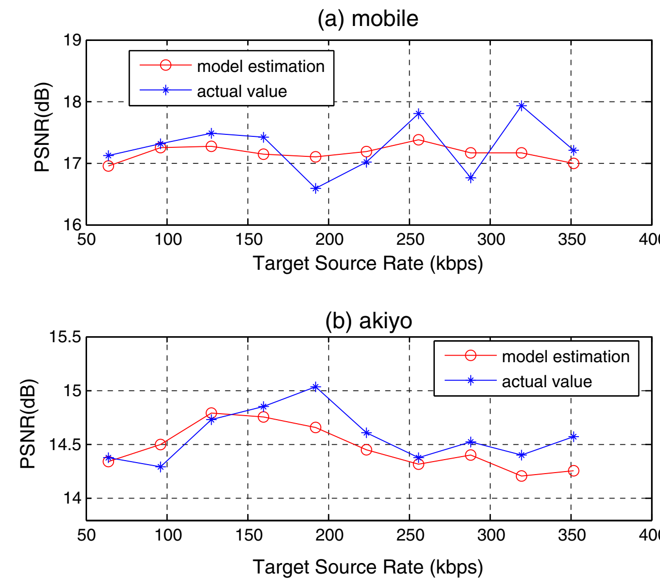


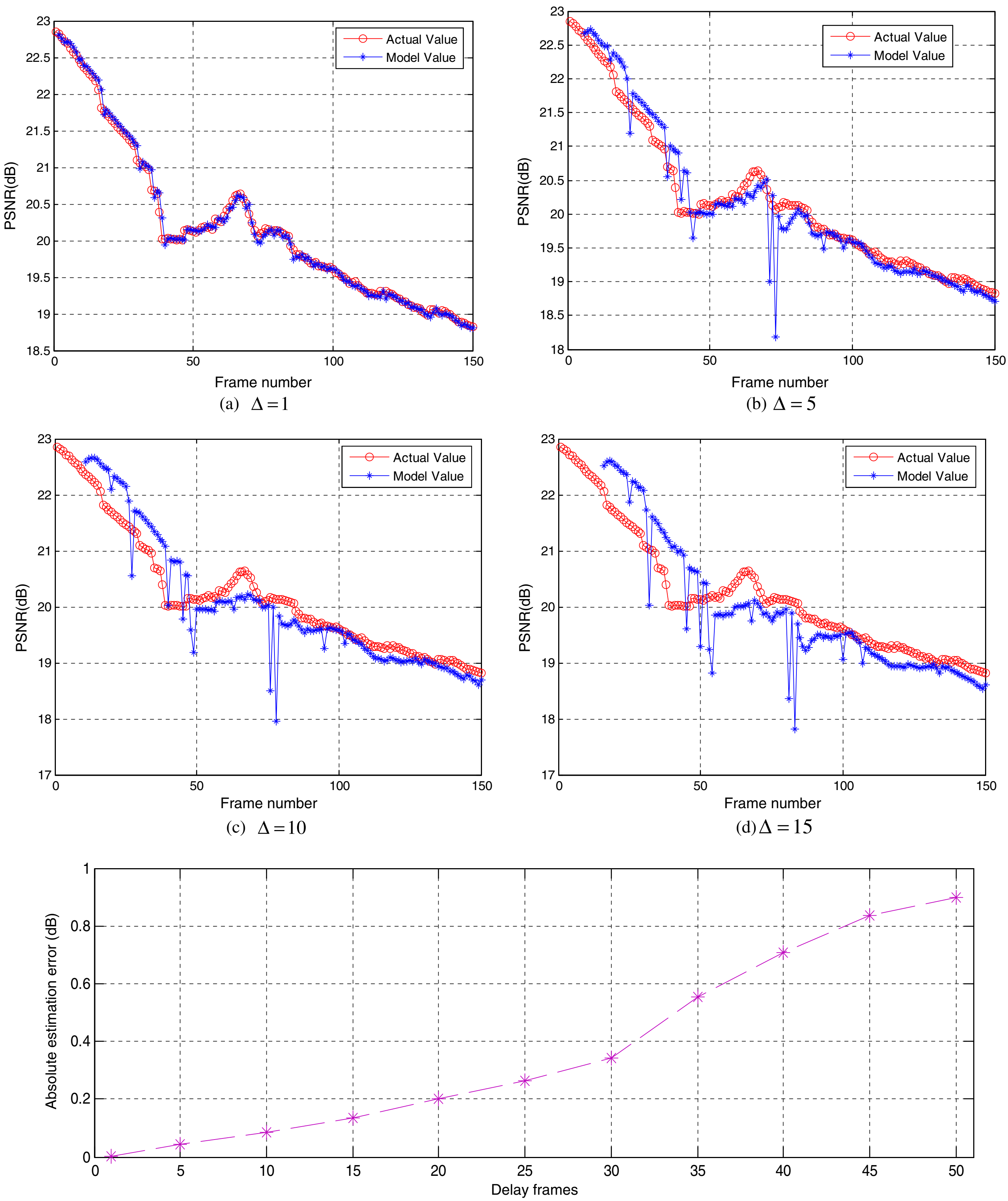
Fig. 3. PSNR performance versus different target source coding rate for estimation.

in the header and marker of a video packet is also integrated into our method unlike the work in [18] which assumed packet headers and markers error free. More importantly, the work in [18] need running many simulations to determine parameters of the model through a video database in off line manner, while our method can work online. We also can see that the major computation is to calculate probabilities and distortion for different cases in the MB level using the proposed method. Compared to our work, [18] need more computation to implement videos simulations to obtain parameters of the model. In addition, there is no an explicit discussion provided by [18] to obtain the model parameters $p(t)$ and $p'_{DC}(t)$, hence, contrast experiments between the work of [18] and ours can not be implemented.

4.2. Simulation results for delayed-feedback channel conditions

In most practical applications, the hybrid channel conditions are time-varying. Applying a feedback channel from the receiver to the sender may provide an effective method to keep the distortion estimation with the change of the channel conditions [7,21]. However, the feedback information such as packet loss status would delay for a time interval to reach the encoder. We assumed that the delayed time interval includes integer number of frames and the channel status is correctly received by the sender. Then the reconstructed quality of the current frame k can be compute recursively from the previous frame $k–\Delta$ by our proposed method, where $\varDelta$ is the delayed frame interval.

We encode the QCIF video *coastguard* with the first VOP as I-VOP followed by 149 P-VOPs. The source coding rate and the frame rate are set 96 kbps and 15 fps, respectively. The estimation and actual results for the delayed frame interval $\varDelta$ = 1, 5, 10 and 15 are shown in Fig. 4. The average packet loss rate and bit error rate are 2.1% and 10^{-5}, respectively. It is noted that here the actual distortion is obtained by running only one simulation. The packet loss ratio and bit error ratio at each frame are shown in Fig. 5(a) and (b), respectively. From the results we can see that when the delayed frame interval is up to 15 frames (equivalently, 1 s for the delay time), the absolute estimation error for a single frame can be maintained quite small. This indicates that the proposed transmission distortion estimation method is very accurate. We can also see that estimation for a certain frame may not catch some actual drastic changes quickly when the delay interval is extended to a large



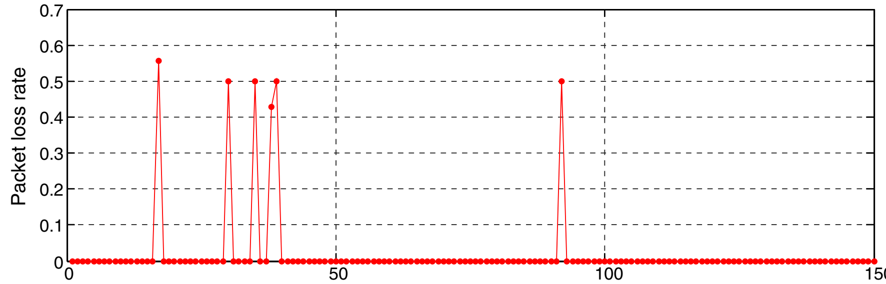
(a) $\Delta = 1$ (b) $\Delta = 5$

(c) $\Delta = 10$ (d) $\Delta = 15$

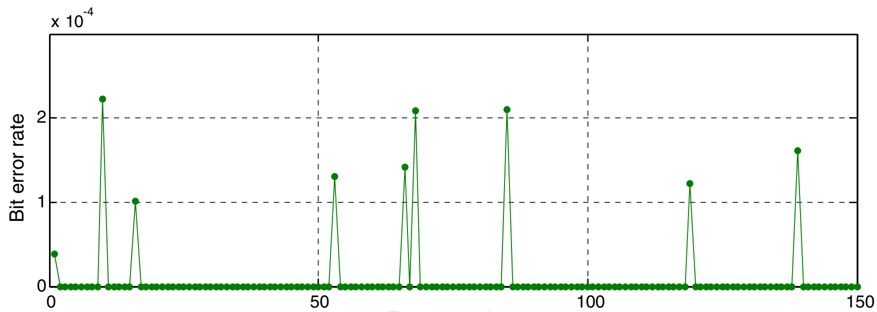
(e) Absolute estimation errors versus different delay frames

Fig. 4. Transmission distortion estimation results for different delay intervals.

scope such as more than 10 frames (equivalently, about 700 ms for 15 frames per second). It is noted that the estimation for the reconstructed quality of the current frame k uses the channel condition information of the previous frame $k-\Delta$, that is, the packet erasure rate and bit error rate of the frame $k-\Delta$ are seen as the same as those of the current frame k. When delayed time becomes larger,



(a) Packet loss rate versus frame number



(b) Bit error rate versus frame number

Fig. 5. The packet loss rate and bit error rate at each frame for time-varying channel.

the channel condition information obtained from the feedback channel may be quite different from the practical information due to time-varying changes of hybrid channels. Thus, the estimated transmission distortion for a single frame tends to deviate from the actual one much more with increasing the delayed interval. Fig. 4(e) shows the absolute estimation error for different delay intervals from 1 frame to 50 frames. The absolute estimation error is an average values over 150 frames. This result indicates that the proposed method is very robust.

5. Conclusion

In this paper, we propose a transmission distortion estimation approach for video delivery over hybrid channels. The method takes into account both packet erasures in wired networks and bit errors in wireless links. It can be used to transmission distortion estimation based on R-D and P-R-D optimization for video delivery over hybrid channels containing wired networks and wireless links such as wireless Internet applications. Applying this method, the sender can estimate the transmission distortion of received decoded video for given or time-varying channel conditions. The experiment results reveal that the method is quite accurate and robust for different videos with different source target coding rates under varying channel configurations. In the future, we would like to apply it to the other video encoders using block-based motion compensation prediction scheme with minor modifications.

Acknowledgements

This work received supports from the National High Technology Research and Development Program of China (863 Program) (Grant No. 2007AA01Z223), the National Natural Science Foundation of China (Grant No. 60572063, 60873127), and Education Ministry Doctoral Research Foundation of China (Grant No. 20070487028).

References

[1] Q. Zhang, Z. Ji, W. Zhu, Y.-Q. Zhang, Power-minimized bit allocation for video communication over wireless channels, IEEE Trans. Circuit Syst. Video Technol. 12 (6) (2002) 398–410.
[2] T.H. Lan, A.H. Tewfik, A resource management strategy in wireless multimedia communications-total power saving in mobile terminals with a guaranteed QoS, IEEE Trans. Multimedia 5 (2) (2003) 267–281.
[3] Zhihai He, Yongfang Liang, Lulin Chen, et al., Power-rate-distortion analysis for wireless video communication under energy constraints, IEEE Trans. Circuit Syst. Video Technol. 15 (5) (2005) 645–658.
[4] Xiaoan Lu, E. Erkip, Yao Wang, et al., Power efficient multimedia communication over wireless channels, IEEE J. Sel. Areas Commun. 21 (10) (2003) 1738–1751.
[5] Y. Eisenberg et al., Joint source coding and transmissionpower management for energy efficient wirelessvideo communications, IEEE Trans. Circuit Syst. Video Technol. 12 (6) (2002) 411–424.
[6] K. Stuhlmuller, N. Farber, M. Link, B. Girod, Analysis of video transmission over lossy channels, IEEE J. Sel. Areas Commun. 18 (6) (2000) 1012–1032.
[7] Zhihai He, Jianfei Cai, Chang Wen Chen, Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding, IEEE Trans. Circuit Syst. Video Technol. 12 (6) (2002) 511–523.
[8] Zhihai He, S.K. Mitra, A unified rate-distortion analysis framework for transform coding, IEEE Trans. Circuit Syst. Video Technol. 11 (12) (2001) 1221–1236.
[9] Hsueh-Ming Hang, Jiann-Jone Chen, Source model for transform video coder and its application. I. Fundamental theory, IEEE Trans. Circuit Syst. Video Technol. 7 (2) (1997) 287–298.
[10] A. Ortega, K. Ramchandran, Rate-distortion methods for image and video compression, IEEE Signal Process. Mag. 15 (6) (1998) 23–50.
[11] R. Zhang, S.L. Regunathan, K. Rose, Video coding with optimal inter/intra-mode switching for packet loss resilience, IEEE J. Sel. Areas Commun. 18 (6) (2000) 966–976.
[12] H. Yang, K. Rose, Recursive end-to-end distortion estimation with model-based cross-correlation approximation, in: Proceedings of International Conference on Image Processing, Barcelona, Spain, 2003, pp. III-469 – III-472.
[13] Yao Wang, Zhenyu Wu, J.M. Boyce, Modeling of transmission-loss-induced distortion in decoded video, IEEE Trans. Circuit Syst. Video Technol. 16 (6) (2006) 716–732.
[14] Zhihai He, Hongkai Xiong, Transmission distortion analysis for real-time video encoding and streaming over wireless networks, IEEE Trans. Circuit Syst. Video Technol. 16 (9) (2006) 1051–1062.
[15] R. Talluri, Error-resilient video coding in the ISO MPEG-4 standard, IEEE Commun. Mag. 36 (6) (1998) 112–119.
[16] Maria G. Martini, Marco Chiani, Proportional unequal error protection for MPEG-4 video transmission, Proceedings of IEEE International Conference on Communications, vol. 4, Helsinki, Finland, 2001, pp. 1033–1037.

[17] M.F. Sabir, H.R. Sheikh, R.W. Heath Jr., A.C. Bovik, A joint source-channel distortion model for JPEG compressed images, IEEE Trans. Image Process 15 (6) (2006) 1349–1364.

[18] M.F. Sabir, R.W. Heath, A.C. Bovik, Joint source-channel distortion modelling for MPEG-4 video, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2006, pp. IV-201 – IV-204.

[19] Y. Pei, J.W. Modestino, Use of concatenated FEC coding for real-time packet video over heterogeneous wired-to-wireless IP networks, in: Proceedings of International Symposium On Circuits and Systems, vol.2, 2003, pp. 25–28.

[20] H. Yousefi'zadeh, H. Jafarkhani, F. Etemadi, Distortion-optimal transmission of progressive images over channels with random bit errors and packet erasures, in: Proceedings of IEEE Data Compression Conference, 2004, pp. 132–141.

[21] Yushi Shen, P.C. Cosman, L.B. Milstein, Video coding with fixed-length Packetization for a tandem channel, IEEE Trans. Image Process. 15 (2) (2006) 273–288.

[22] F. Etemadi, H. Jafarkhani, A low complexity progressive bitstream transmission system for hybrid channels with correlated loss, in: Proceedings of IEEE International Conference on Image Processing, 2006, pp. 1–4.

[23] GuiJin Wang, Qian Zhang, Wenwu Zhu, Channel-adaptive error protection for scalable video over channels with bit errors and packet erasures, in: IEEE International Symposium on Circuits and Systems, 2002, pp. II-712 – II-715.

[24] ISO-IEC/JTC1/SC29/WG11/N2202, Information technology-coding of audio-visual objects: visual – ISO/IEC 14496-2 Committee Draft, March, 1998.

[25] Y.Q. Shi, Xi Min Zhang, Zhi-Cheng Ni, et al., Interleaving for combating bursts of errors, IEEE Circuits Syst. Mag. 4 (1) (2004) 29–42.

[26] Jae-Young Pyun, Jae-Jeong Shim, Sung-Jea Ko, et al., Packet loss resilience for video stream over the Internet, IEEE Trans. Consumer Electron. 48 (3) (2002) 556–563.