Available online at www.sciencedirect.com
SCIENCE DIRECT®
J. Vis. Commun. Image R. 16 (2005) 475–498
JOURNAL OF VISUAL Communication & IMAGE Representation
www.elsevier.com/locate/jvci

Packet video transmission over wireless channels with adaptive channel rate allocation ☆

Wei-Ying Kung [a], Chang-Su Kim [b], C.-C. Jay Kuo [a,*]

[a] *Integrated Media Systems Center, Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 2564, USA*
[b] *Department of Information Engineering, The Chinese University of Hong Kong Shatin, NT, Hong Kong*

Received 28 March 2004; accepted 11 March 2005
Available online 20 April 2005

Abstract

A robust video communication system based on layered coding and unequal error protection is developed in this work. We consider two video communication scenarios. First, for pre-compressed video bitstreams, a channel code rate allocation scheme is proposed to minimize the expected mean square error subject to a constraint on the overall bit budget. Second, for real-time video transmission, we jointly optimize the quantization parameters and the channel coding rates according to channel conditions. To this end, we develop a simple rate-distortion model for general video coders using DCT and motion compensation, so that the rate and the distortion can be estimated without an expensive encoding procedure. Simulation results show that the proposed algorithms provide acceptable image quality even in high bit error rate environments.

Keywords: Packet video; Wireless video; Robust video transmission; Joint source-channel coding; Unequal error protection

☆ The research has been funded in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, Cooperative Agreement No. EEC-9529152.
* Corresponding author. Fax: +1 213 740 4651.
E-mail addresses: weiyingk@ieee.org (W.-Y. Kung), cskim@ieee.org (C.-S. Kim), cckuo@sipi.usc.edu (C.-C. Jay Kuo).

1047-3203/$ - see front matter
doi:10.1016/j.jvcir.2005.03.005

1. Introduction

With the increasing demand on multimedia services and the stimulus of 4th generation wireless systems, it has drawn much attention from industry and academia to design a suitable method for transmitting multimedia signals over wireless channels. However, wireless channels contain various types of noises and the link quality can be severely degraded due to shadowing and fading effects. Multimedia signals, such as audio, image and video, are susceptible to transmission errors. It is necessary to design a robust transmission scheme to protect the quality of multimedia signals against transmission errors.

Raw video data require a huge amount of bandwidth for their storage and transmission. Attempts have been made to compress video data efficiently and several international standards, such as MPEG-4 [1] and H.263 [2], have been proposed for the compression of digital video. These standards achieve a high compression ratio by exploiting temporal and spatial correlations in an image sequence with motion compensation (MC) and the discrete cosine transform (DCT). But, as the image sequence is more highly compressed, the encoded bitstream becomes more vulnerable to transmission errors.

Table 1 compares four strategies to enhance the error resilience of video bitstreams against transmission errors. They are error protection, error confinement, error concealment, and interactive error control. The error protection strategy inserts redundant information into compressed data so that the decoder can recover from transmission errors using the redundant information. The error confinement reduces the use of predictive encoding operations to localize the effect of transmission errors. Both the error protection and the error confinement are effective especially in high bit error rate (or packet loss rate) environments. But, they sacrifice coding efficiency and cannot guarantee an error-free transmission. Thus, the error concealment strategy is employed at the decoder side to hide the effect of remaining errors. The error concealment neither introduces extra delay nor lowers the compression ratio, but its error resilience capability is limited. Finally, the interactive error control strategy requires a feedback channel to adjust the encoding mode according to channel conditions or acknowledge signals from the decoder. This strategy accomplishes error resilience in the most efficient way, but requires extra delay and may not be suitable in multicast and broadcast applications. To summarize, no strategy always performs better than the others. To select a suitable resilient method, we should consider application requirements such as coding efficiency, delay tolerance, and implementation complexity.

We focus on the error protection strategy in this work. From the information theoretical viewpoint, the optimum reliable transmission can be achieved by separating

Table 1
Comparison of the strategies to improve error resilience

	Error protection	Error confinement	Error concealment	Interactive error control
Coding efficiency	Poor	Fair	Excellent	Good
Extra delay	No	No	No	Yes
Error resilience	Excellent	Good	Fair	Excellent

the source coding from the channel coding [3]. However, in real applications under finite delay and complexity constraints, we can improve the quality of reconstructed video by jointly designing the source and the channel coders. Video signal is a kind of distortion tolerant media and human eyes are not equally sensitive to all bits in a video bitstream. It is thus advantageous to partition compressed video data into several layers so that they can be treated in different ways according to their importance levels. A network, which is capable of providing different service levels (DiffServ), can transmit each layer at its corresponding priority.

However, many wireless and Internet systems are not intelligent enough to support DiffServ. In such a case, we can provide unequal protection by employing forward error correction (FEC) codes, such as block codes and convolutional codes [4–8]. Rate-compatible punctured convolutional (RCPC) codes are especially suitable for unequal error protection, since they can flexibly adapt the level of protection by changing puncturing tables [9]. An alternative approach to provide DiffServ is to use the different bit error rates, caused by the characteristics of signal constellations in modulation [10,11].

Due to the limited bandwidth and the delay sensitivity of real-time video transmission, the optimal trade-off between source and channel coding under rate constraint [12–15], delay constraint [16] or power constraint [17,18] has been studied extensively [19]. A typical joint source/channel coder [12–15] is designed to adaptively choose the quantizer of source coder and the channel rate of FEC coder. When the channel becomes very noisy, a coarser quantizer and a lower rate FEC code are selected to combat channel errors. In contrast, when the channel condition becomes good, a finer quantizer and a higher rate FEC code can be selected to transmit high quality video. It is not feasible to compute all possible combination of quantizer step size and channel rate to find the best operating point. Thus, many attempts have been made to develop a rate-distortion model for video signals to reduce the computational complexity [20–27].

In this work, we first develop a video coder based on layered coding and interleaved packetization. Then, we consider two scenarios of video communications. The first scenario is to protect the packets of a pre-compressed video signal, when the channel condition is fixed. We propose a channel code rate allocation scheme to minimize the expected mean square error subject to a constraint on the overall bit rate. In the second scenario, we jointly allocate the source and the channel rates to each packet in real-time video transmission over a wireless channel, whose bit error rate is fluctuating. We develop a simple rate-distortion model, so that the rate and the distortion can be estimated without an extensive encoding procedure. Simulation results demonstrate that the proposed algorithms for both scenarios provide acceptable image quality even in high bit error rate environments.

2. Video coder

We employ a video coder, which is modified from the standard H.263 coder [2]. In the encoder, we partition compressed video data into base and enhancement layers and packetize them in an interleaved way to enhance error resilience. In the decoder, we use a motion-compensated error concealment scheme to recover corrupted video regions faithfully.

2.1. Layered coding and packetization

In the encoder, we first partition compressed video data into two layers as shown in Fig. 1. The base layer contains macroblock (MB) headers and motion vectors, whose loss degrades the received video quality severely. The enhancement layer contains less important information, i.e., the residual DCT coefficients after motion compensation. For an intra MB, DCT coefficients are put into the base layer.

We then perform the packetization at the base and the enhancement layers, respectively. Due to the packetization, the effect of errors can be localized within a packet. As the packet size becomes smaller, the error localization becomes more effective. However, a smaller packet size introduces more overhead bits to distinguish separated data parts. Furthermore, a larger amount of computations are required to evaluate the importance of each packet so that packets can be treated differently according to their importance. Thus, the bit rate overhead, the computational complexity as well as the error localization capability should be taken into account to determine the packet size.

In H.263, a synchronization code can be inserted at the group of blocks (GOB) level. Similarly, we can partition the encoded video stream into GOBs. In this way, a QCIF (176×144) frame can be partitioned to 9 base packets and 9 enhancement packets. However, if a GOB is corrupted, only the upper and lower MBs are available to conceal an erroneous MB.

In this work, we introduce an alternative way to reorganize MBs so that each packet consists of sparsely distributed MBs. Thus, the decoder can conceal the erroneous MB more effectively by using the information of more neighboring MBs. Fig. 2 illustrates the proposed packetization scheme. A packet for the QCIF format video is formed with 11 MBs chosen from every nine consecutive MBs. Specifically, 1st, 10th, 19th, and 28th MBs are grouped into one packet, 2nd, 11th, 20th, and 29th MBs are grouped into another packet, and so on. Therefore, as in the GOB packetization, the proposed scheme also generates 9 base packets and 9 enhancement packets for each frame. But, when a packet is missing, the erroneous MBs can be concealed by using the information of the upper, lower, left and right MBs. At the end of each packet, a 16-bit cyclic redundancy code (CRC) is added to enable the error detection at the decoder side.

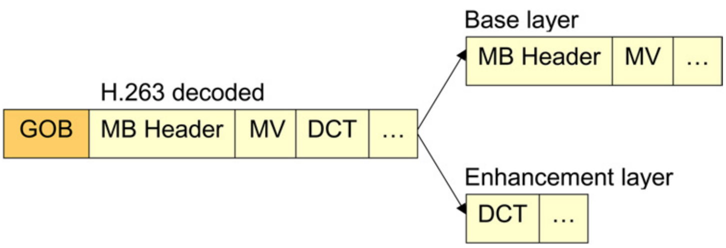

Fig. 1. Illustration of data partitioning.

1	2	3	4	5	6	7	8	9	1	2
3	4	5	6	7	8	9	1	2	3	4
5	6	7	8	9	1	2	3	4	5	6
7	8	9	1	2	3	4	5	6	7	8
9	1	2	3	4	5	6	7	8	9	1
2	3	4	5	6	7	8	9	1	2	3
4	5	6	7	8	9	1	2	3	4	5
6	7	8	9	1	2	3	4	5	6	7
8	9	1	2	3	4	5	6	7	8	9

Fig. 2. MB grouping for a QCIF frame.

2.2. Error concealment

Error protection schemes can reduce bit error rate, but they cannot guarantee to correct all bit errors. Since compressed video data consist of variable length codewords, they are vulnerable to even a single bit error. Using the CRC code at the end of each packet, the decoder first performs the error detection. If a packet is detected as erroneous, the decoder discards all the data within the packet and conceals the loss.

If an enhancement packet is lost but the corresponding base packet is intact, we replace the missing DCT coefficients with zeros and copy each MB from the previous frame using the motion vectors in the base packet. This approach provides a good image quality, since we can exploit high temporal correlation in image sequences with correct motion vectors.

On the other hand, if a base packet is lost, the corresponding enhancement packet is useless. A simple approach is to directly copy the missing MBs from the previous frame with zero motion vector. This approach gives an acceptable performance when a sequence contains only slow motions. But, in a fast moving sequence, the direct copying results in obvious discontinuities and artifacts.

In this work, the loss of a base packet is concealed in the following way. Suppose that a corrupted MB is surrounded by four correctly received inter MBs. For each missing pixel p in the corrupted MB, four predicted pixels are obtained from the previous frame by the motion vectors of the upper, lower, left and right MBs, respectively. They are denoted by p_{upper}, p_{lower}, p_{left}, and p_{right}. To conceal the pixel p, the four predicted pixel values are averaged using weighting coefficients, which are inversely proportional to the distances between p and the adjacent MBs. Specifically, assume that p is the (x, y)th pixel in the missing MB, where $1 \leqslant x, y \leqslant 16$. Then, it is concealed via

$$\tilde{p} = \frac{p_{\mathrm{upper}} \cdot (17 - y) + p_{\mathrm{lower}} \cdot y + p_{\mathrm{left}} \cdot (17 - x) + p_{\mathrm{right}} \cdot x}{34}.$$

Note that the weighting coefficient for p_{left} is $\frac{17-x}{34}$. It approaches 0.5, as p becomes closer to the left MB (i.e., x is smaller). In contrast, it approaches 0, as p becomes

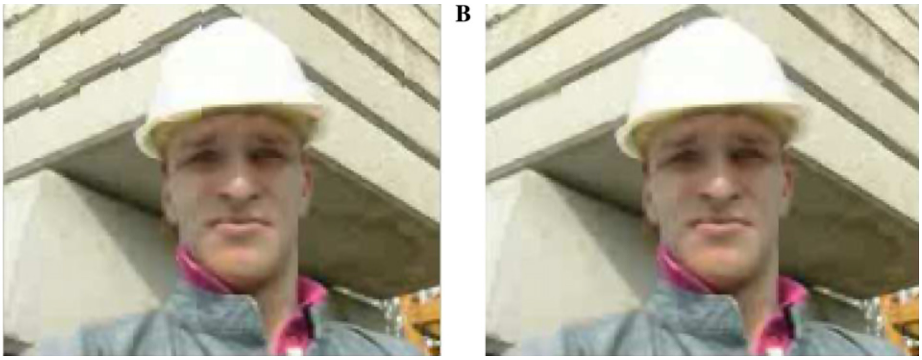

Fig. 3. Loss of the 3rd base packet. (A) The erroneous region is copied from the previous frame. (B) The erroneous region is concealed by the proposed algorithm.

closer to the right MB. The weighting coefficients for the other predicted pixels are defined in a similar way. If a neighboring motion vector is not available due to the packet loss, intra coding mode or boundary effect, only those available motion vectors are used for the concealment. If all motion vectors are not available, the erroneous MB is copied from the previous frame with zero motion vector.

Fig. 3 compares the reconstructed frames of the direct copying method and the proposed algorithm, when the 3rd base packet for the 2nd frame of 'Foreman' sequence is lost. We see clearly that the copying algorithm yields obvious discontinuities in the background building, while the proposed algorithm recovers the motion vectors smoothly and provides very faithful image reconstruction without noticeable artifacts.

3. Static channel code rate allocation

3.1. RCPC

Error correction codes are used to combat undesired noises and interferences in wireless environments. Since our goal is to vary the channel coding rate according to different QoS requirements, we adopt the rate-compatible punctured convolutional (RCPC) code [9] due to its flexibility in adjusting the rate. A convolutional code with a coding rate of k/n can be represented with n linear algebraic function generators and implemented by passing an input sequence through a finite state machine with k-bits per stage. As shown in Fig. 4, in RCPC coding, certain output bits are deleted according to the puncturing table after the convolutional encoding to increase the channel coding rate. At the decoder side, '0's are inserted using the same puncturing table as in the encoder before the Viterbi decoding. Efficient generator functions and tables can be found in related textbooks and papers [9].

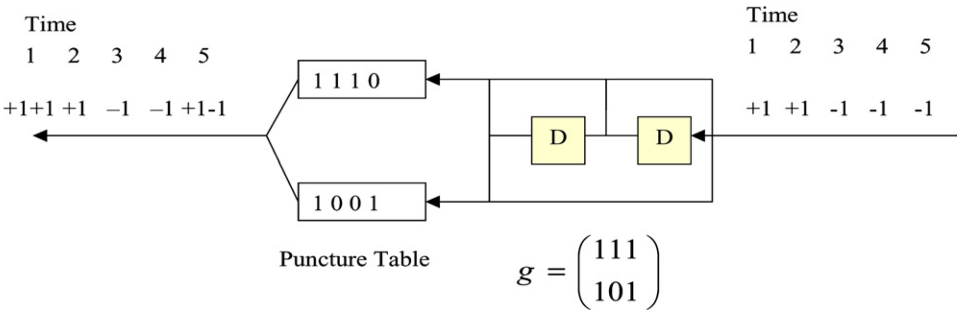

Fig. 4. RCPC with rate = 4/5.

3.2. Channel distortion

It is necessary to assign the channel code rate to each packet according to its importance so that the limited bandwidth can be used efficiently. An ideal way to measure the packet importance is to associate each packet with the effect of its loss on human perception. However, human eyes are very complicated organs, and it is difficult to obtain the exact relation between video packet and its perceptual quality. Thus, for simplicity, we use the mean square error (MSE) to measure packet loss effect. The MSE measurement of each packet loss is straightforward. We emulate the dropping of each packet, perform its error concealment and then compute MSE between the concealed data and the error-free data.

Let δ_i^2 denote the MSE value due to the loss of packet i. Then the expected channel distortion for a frame can be written as

$$D_{\text{channel}} = \sum_{i=1}^{n} \delta_i^2 \cdot P_i, \tag{1}$$

where n is the number of packets for the frame and P_i is the probability that packet i is lost. The packet loss rate P_i is given by

$$P_i = 1 - (1 - e)^{N_i}, \tag{2}$$

where e is the bit error rate when a certain RCPC channel code rate is applied, and N_i is the size of packet i in bits. We assume that even a single bit error ruins the whole packet. In other words, an uncorrupted packet means that it contains no bit error after the RCPC decoding.

The objective is to determine the RCPC channel code rate for each packet, which minimizes the distortion in Eq. (1) subject to a certain rate constraint. Therefore, we need to find the relation between the channel code rate and the bit error rate e in Eq. (2). Since it is difficult to find a closed form relation, we use several training sequences to measure the bit error rate when a certain channel code rate is used. In this way, we can find an experimental mapping of channel code rate and bit error rate. Then, we can solve the problem by the following exhaustive search:

(1) Get the MSE value δ_i^2 for each packet in a frame.
(2) Try every combination of channel code rates, and select the set that minimizes the expected distortion in Eq. (1) subject to the constraint on the overall transmission rate.
(3) Apply RCPC to each packet with the selected channel code rate.

3.3. Fast search algorithm

The rate assignment method described above is computationally expensive. In our implementation, there are four choices of channel code rates (8/8, 8/16, 8/24 and 8/32), and we process a frame as a unit for the assignment. For example, in a QCIF frame, there are 18 packets, and the exhaustive search should check 4^{18} candidates to find the best channel rate combination. Such a high complexity is unacceptable, and a fast search algorithm is necessary in real time applications.

The Lagrangian multiplier method [19] is often employed to solve the rate-distortion optimization problem, though it has the limitation that it can only reach points on the convex hull of the achievable R-D region. Unfortunately, in our case, the convex hull is not dense enough and the Lagrangian solution may not provide a reasonably good performance. Although the dynamic programming approach can overcome this problem, its complexity increases exponentially with both the number of packets and the number of rate choices.

We develop a different fast search method. As given in the cost function in Eq. (1), there are two factors that affect the assignment result. One is the MSE value of each packet loss, and the other is the packet loss rate that is a function of the packet size. If we assume that all packets have the same size, the assignment depends only on the MSE value. Specifically, the packet with a higher MSE value requires stronger protection. Thus, if we arrange packets in the descending order of the MSE values, the assigned rates for those packets should be in the ascending order. The assigned channel code rates also should satisfy the overall transmission rate constraint. But, if the total rate is much lower than the rate constraint, we can lower the channel code rates of some packets to provide stronger protection. Thus, the best channel rate combination happens only when the total transmission rate is close to the rate constraint. The following procedure is developed based on these observations:

(1) List the packets in the descending order of their MSE values.
(2) Find the set S of the channel code rate vectors $\mathbf{r}_k$ that satisfies the ascending rule. Specifically, $S = \{\mathbf{r}_k = (r_{k,1}, r_{k,2}, \ldots, r_{k,n}) : r_{k,i} \leqslant r_{k,j} \text{ for } i < j\}$, where $r_{k,i}$ denotes a channel code rate for packet i.
(3) Find the subset S' of S that contains only the vectors satisfying the overall transmission rate constraint.
(4) Let $\mathbf{r}_l = (r_{l,1}, r_{l,2}, \ldots, r_{l,n})$ and $\mathbf{r}_m = (r_{m,1}, r_{m,2}, \ldots, r_{m,n})$ be two elements in S'. Note that the protection capability of $\mathbf{r}_l$ is inferior to that of $\mathbf{r}_m$, if $r_{l,i} \geqslant r_{m,i}$ for all i. Thus, we remove such elements from S' by comparing every pair of elements in S'. Let S'' denote the resulting subset of S'.

(5) Select the suboptimal code rate vector from S'' that minimizes the cost function in Eq. (1).

In the above procedure, we reduce the candidate set from S to S''. In general, the size of S'' is much smaller than that of S, reducing the search time significantly. For example, let us assume that there are 6 packets and each packet can have one of the four channel rates 8/8, 8/16, 8/24, and 8/32. When the overall transmission rate is constrained to be smaller than twice the original source rate, the above procedure yields S'' consisting of only 7 candidate vectors in Table 2. On the contrary, if the full search method is employed, we need to calculate the cost function $4^6 = 4096$ times.

Although the same packet size assumption was made in developing the fast search algorithm, packet sizes are actually varying in our system. It is, however, worthwhile to point out that packet sizes do not vary significantly, since we employ a rate control scheme at the video encoder. Experimental results confirm that the fast search algorithm provides an acceptable performance at a low complexity. For a QCIF frame, the fast search algorithm usually takes around 40 time units to find the solution instead of 4^{18} time units for the full search. Table 3 compares the search results of the Lagrangian multiplier method and the fast search algorithm. We see that the fast search algorithm tends to achieve a significantly lower cost than the Lagrange multiplier method.

Table 2
An example of the fast channel code rate assignment

	Packet 1	Packet 2	Packet 3	Packet 4	Packet 5	Packet 6
1	8/16	8/16	8/16	8/16	8/16	8/16
2	8/24	8/16	8/16	8/16	8/16	8/8
3	8/24	8/24	8/16	8/16	8/8	8/8
4	8/24	8/24	8/24	8/8	8/8	8/8
5	8/32	8/16	8/16	8/16	8/8	8/8
6	8/32	8/24	8/16	8/8	8/8	8/8
7	8/32	8/32	8/8	8/8	8/8	8/8

Table 3
Comparison of different search results

	Lagrangian multiplier method	Fast search algorithm
Frame 1	Cost = 18.699 Rate = 1518	Cost = 6.27394 Rate = 1478
Frame 2	Cost = 40.6726 Rate = 1741	Cost = 7.03531 Rate = 1741
Frame 3	Cost = 31.4817 Rate = 1652	Cost = 6.7466 Rate = 1927

4. Adaptive source and channel rate allocation

In the previous section, we developed a channel rate allocation scheme for the case when the source bit rate and the channel condition are fixed. However, the channel condition varies dynamically in wireless video communications. In such cases, it is advantageous to jointly allocate the source and the channel rates according to the fluctuating channel condition.

4.1. Video source model

A typical video encoder performs motion estimation, DCT, quantization, and variable length coding (VLC). These components affect one another. For example, if the motion estimation module can find a well-matched block in the previous frame or the quantization is performed with a large step size, VLC requires a small amount of bits to encode residual DCT coefficients. Also, the distribution of residual coefficients highly depends on the characteristics of the input image sequence. It is hence not easy to develop a statistic model that precisely predicts bit rates and distortions. However, in this work, we attempt to develop a simple model that can estimate the bit rate and the distortion of a video packet in terms of the quantization parameter Q.

4.1.1. Bit rate model

The bit rate can be approximated as the entropy of quantized coefficients [28]. However, the empirical rate is usually lower than the 0th-order entropy, which is computed by assuming that the coefficients are independent of one another. This is because the coefficients are encoded by the run-length coding, exploiting consecutive zero coefficients. The discrepancy between the entropy and the empirical rate becomes larger as the image sequence is encoded at a lower bit rate with a larger Q. Since we focus on the low bit rate coding for wireless applications, the entropy method is not suitable in our approach. Recently, another method called ρ-domain R-D analysis was proposed in [25], where the relationship between the bit rate and the percentage ρ of zeros in quantized DCT coefficients was analyzed. However, to achieve an accurate rate estimation, the method proposed in [25] requires a computationally expensive process to obtain about 10 model parameters.

To maintain a low computational complexity, we attempt to find a bit rate function in terms of Q rather than through several intermediate parameters. Fig. 5 shows the average bit rate for three sequences 'Foreman,' 'Claire' and 'Salesman.' Let us analyze the bit rates for base packets and enhancement packets separately.

Enhancement packets contain residual DCT coefficients, and their bit rate decreases hyperbolically as Q increases. In Fig. 5, we also plot the polynomial function

$$R_{\mathrm{enh}} = \frac{A}{Q^2} + B,$$

which approximates the enhancement bit rate. Parameters A and B are obtained to minimize the mean square approximation error. It is observed that the polynomial

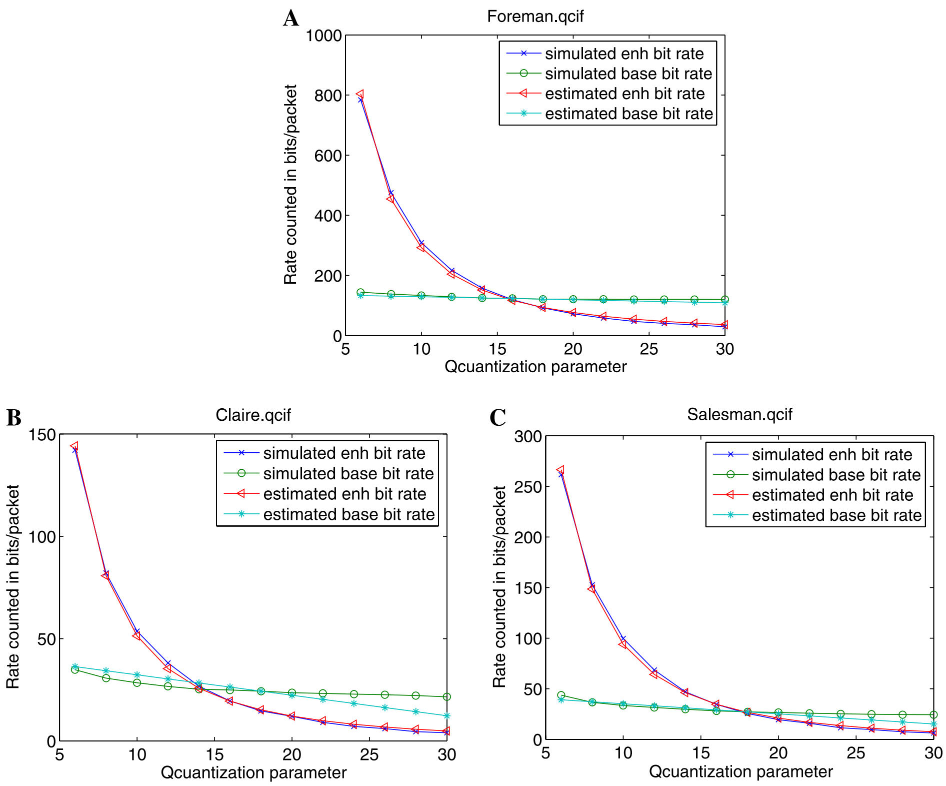

Fig. 5. The rate-quantization curves for (A) 'Foreman,' (B) 'Claire,' and (C) 'Salesman' sequences.

functions approximate very well the empirical bit rates. Table 4 summarizes parameters A and B for the three sequences. Note that B is negligible as compared to A. Thus, we can approximate the bit rate further with

$$R_{\text{enh}} = \frac{A}{Q^2}. \tag{3}$$

Parameter A depends on source characteristics. 'Foreman' sequence contains the fastest motion, thus its DCT coefficients are more widely distributed and require a higher bit rate than those of the other two sequences. Thus,'Foreman' sequence

Table 4
Parameters for the rate-quantization model $R_e = \frac{A}{Q^2} + B$

	A	B
Foreman	28778	5
Claire	5226.4	−0.9
Salesman	9713.5	−3.3

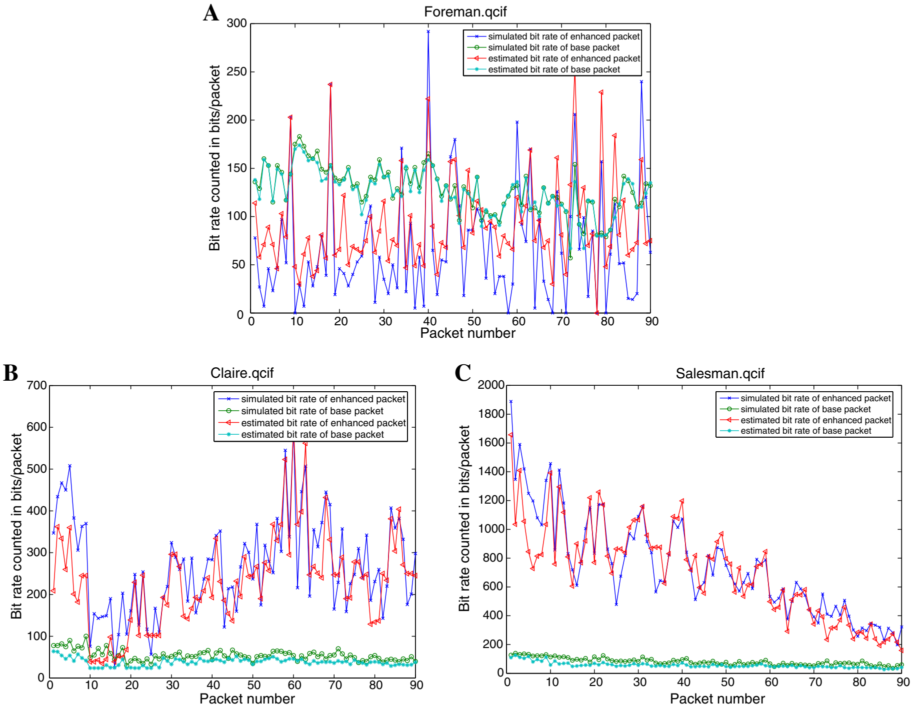

Fig. 6. The bit rate estimation for (A) 'Foreman,' (B) 'Claire,' and (C) 'Salesman' sequences.

has the largest A in Table 4. Parameter A is estimated by encoding a packet with a sample quantization parameter Q_s. We obtain the resulting bit rate $R_{\mathrm{enh,s}}$ for the packet. Then, A is estimated as $A = R_{\mathrm{enh,s}} \times Q_s^2$. Finally, Eq. (3) is used to estimate the bit rates of the enhancement packet at the other quantization parameters.

On the other hand, base packets contain MB headers and motion vectors, thus their bit rate does not vary significantly as Q changes. It is observed that the average bit rate for base packets linearly decreases as Q increases, where the slope is empirically found to be around 1. Thus, the bit rate for base packets can be approximated via

$$R_{\mathrm{base}} = R_{\mathrm{base,s}} + (Q_s - Q), \tag{4}$$

where $R_{\mathrm{base,s}}$ is the bit rate when Q_s is selected as the sample quantization parameter.

Fig. 6 compares the empirical bit rates and the estimated bit rates. It can be seen that the estimated bit rates for both base and enhancement packets are very close to the empirical ones.

4.1.2. Quantization distortion model

In a typical video coder, only the quantization of DCT coefficients incurs the source distortion D_{source}, whereas the other components process video signals losslessly. We adopt the mean square error as the measure for the quantization distortion, which is given by

$$D_{\text{source}} = E[(f - \hat{f})^2] = \int_{a_L}^{a_U} (f - \hat{f})^2 p(f)\, \mathrm{d}f,$$

where f is a DCT coefficient ranging from a_L to a_U, $\hat{f}$ is the quantized output, and $p(\cdot)$ is the probability density function of f. The distribution of DCT coefficients is often modelled as the Laplacian distribution [29], given by

$$p(x) = \frac{\mu}{2} \exp^{-\mu|x|}. \tag{5}$$

Then, we have the quantization distortion

$$D_{\text{source}} = \sum_{i=-\infty}^{\infty} \int_{q(i-1/2)}^{q(i+1/2)} (x - qi)^2 p(x)\, \mathrm{d}x,$$

where q denotes the quantization step size and i denotes the quantization index. Note that the quantization parameter Q is half the step size in H.263 (i.e., $q = 2Q$). As shown in Appendix A, the quantization distortion can be approximated as

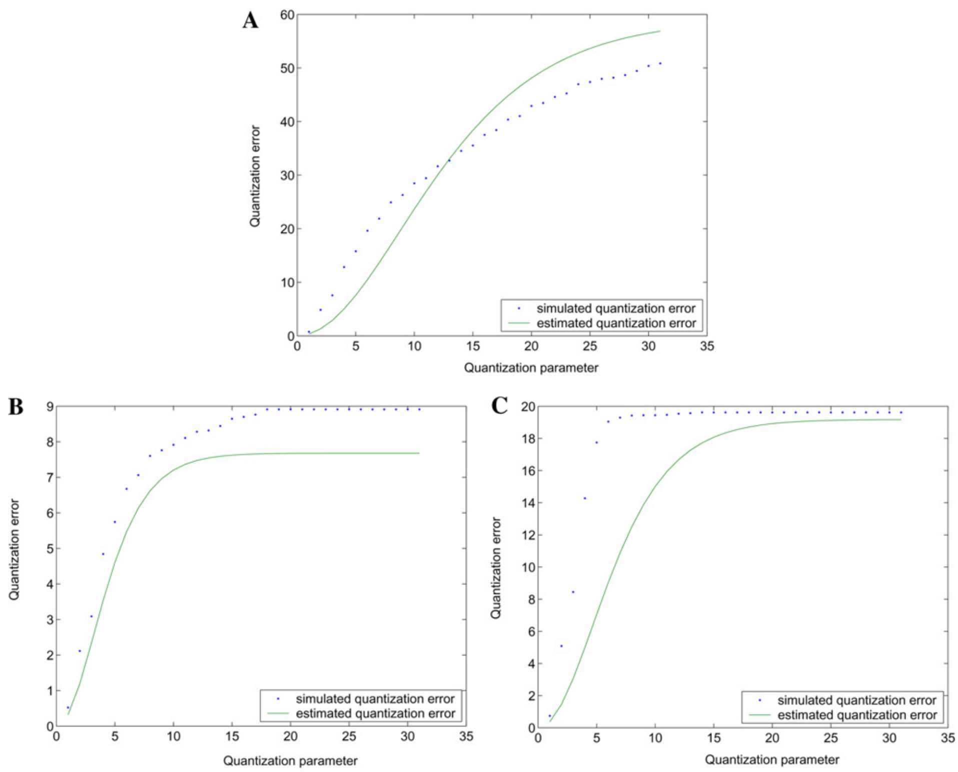

Fig. 7. The distortion-Q curves for (A) 'Foreman,' (B) 'Claire,' and (C) 'Salesman' sequences.

$$D_{\text{source}} = \frac{2}{\mu^2}\left[1 - \frac{e^{-1/2\mu q} - e^{-128.5\mu q}}{1 - e^{-\mu q}}(\mu q)\right] \approx \frac{2}{\mu^2}\left[1 - \frac{e^{-\mu Q}}{1 - e^{-2\mu Q}}(2\mu Q)\right]. \tag{6}$$

Fig. 7 plots the quantization distortion in terms of the quantization parameter Q. At the same Q, 'Foreman' sequence has the highest distortion, since it contains fast movements. The distortions of 'Claire' and 'Salesman' reach saturation points when Q is greater than 10, i.e., when most DCT coefficients are quantized to zeros. The proposed model tends to underestimate D_{source}, when a small value of Q is chosen. This is because there is a large discrepancy between the actual distribution and the Laplacian distribution at a small Q. Nevertheless, the proposed model provides a much more accurate estimation of D_{source} than the traditional model

$$D_{\text{source}} = \frac{q^2}{12} = \frac{Q^2}{3}$$

that assumes the uniform distribution of coefficients.

For each packet, after DCT but before the quantization, the Laplacian parameter μ can be obtained from the variance σ^2 of DCT coefficients by

$$\mu = \sqrt{\frac{2}{\sigma^2}}.$$

Then, D_{source} can be estimated from the quantization parameter Q via Eq. (6). Fig. 8 shows that the proposed model effectively estimates the quantization distortion for each packet, even though the distribution of DCT coefficients varies dynamically according to the motion and the texture of video contents.

4.2. Channel distortion

In the proposed adaptive source and channel rate allocation, a GOB is a basic adaptation unit. A GOB consists of a base packet and an enhancement packet. Similar to Eq. (1), the channel distortion due to possible losses of the base and the enhancement packets can be written as

$$D_{\text{channel}} = \delta^2_{\text{base}} \cdot P_{\text{base}} + \delta^2_{\text{enh}} \cdot P_{\text{enh}}. \tag{7}$$

The MSE values δ^2_{base} and δ^2_{enh} are computed by emulating the dropping and concealment of the base and the enhancement packets, respectively. Also, the packet loss rates P_{base} and P_{enh} can be computed from the packet sizes via Eq. (2).

The quantization parameter Q also affects δ^2_{base} and δ^2_{enh}, and its influence on the channel distortion should be investigated. An enhancement packet consists of residual DCT coefficients only. When it is lost, all coefficients are set to zeros. If we assume that there is no quantization distortion, δ^2_{enh} is equal to the variance σ^2 of the residual coefficients. In general, it was found experimentally that δ^2_{enh} is inversely proportional to the quantization distortion D_{source} and can be approximated as

$$\delta^2_{\text{enh}} = \sigma^2 - D_{\text{source}}. \tag{8}$$

Also, the motion-compensated error concealment is applied to a missing base packet. From definition, δ^2_{base} is the mean square difference between the concealed recon-

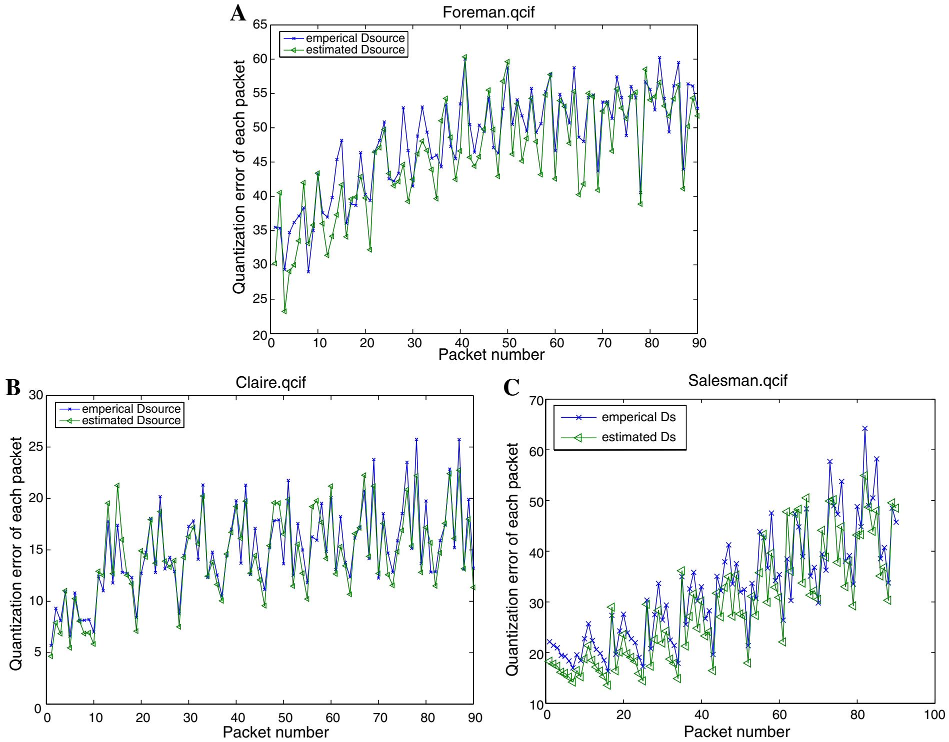

Fig. 8. The estimation of the quantization distortion D_{source} for each packet: (A) 'Foreman,' (B) 'Claire,' and (C) 'Salesman' sequences.

struction and the error-free reconstruction. The concealed reconstruction is based on the previously reconstructed frame and the neighboring motion vectors so that it is independent of the current Q. But, the error-free reconstruction is dependent on Q. We calculate the mean square difference, $\delta^2(Q_s, Q)$, between the reconstructed frames using the quantization parameters Q_s and Q. Then, δ^2_{base} is approximated by

$$\delta^2_{\text{base}} = \delta^2_{\text{base,s}} - \delta^2(Q_s, Q), \tag{9}$$

where $\delta^2_{\text{base,s}}$ is the mean square error due to the loss of the base packet when the sample quantization parameter Q_s is employed. The above formula indicates that δ^2_{base} decreases as Q increases, which is consistent with empirical results.

Fig. 9 shows that the estimated distortions are close to the empirical data for 'Foreman,' 'Claire,' and 'Salesman' sequences.

4.3. Adaptation

The overall distortion D is the sum of the source and the channel distortions. From Eqs. (6)–(9), we have

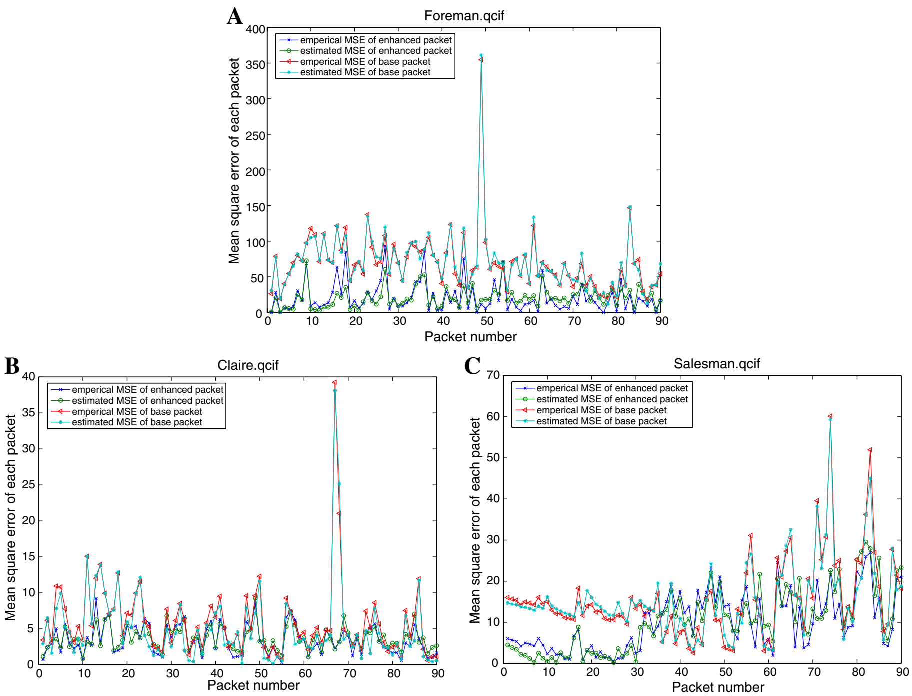

Fig. 9. The estimated distortion due to each packet loss: (A) 'Foreman,' (B) 'Claire,' and (C) 'Salesman' sequences.

$$D = D_{\text{source}} + D_{\text{channel}} = [\delta^2_{\text{base,s}} - \delta^2(Q_s, Q)]P_{\text{base}} + \sigma^2 P_{\text{enh}} + \frac{2}{\mu^2}\left[1 - \frac{e^{-\mu Q}}{1 - e^{-2\mu Q}}(2\mu Q)\right](1 - P_{\text{enh}}), \tag{10}$$

where σ^2 is the variance of DCT coefficients and $\mu = \sqrt{\frac{2}{\sigma^2}}$.

Before the RCPC coding, the bit rates of the base and the enhancement packets are given by Eqs. (4) and (3), respectively. Then, the overall transmission rate can be written as

$$R = \frac{R_{\text{base}}}{C_{\text{base}}} + \frac{R_{\text{enh}}}{C_{\text{enh}}} = \frac{R_{\text{base,s}} + Q_s - Q}{C_{\text{base}}} + \frac{A/Q^2}{C_{\text{enh}}}, \tag{11}$$

where C_{base} and C_{enh} are the RCPC channel code rates for the base and the enhancement packets.

Our goal is to optimize the quantization parameter and the channel code rates for each pair of base and enhancement packets, which minimize the overall distortion D in Eq. (10) subject to the constraint that the overall rate R in Eq. (11) is lower than a certain bit rate. The optimization is performed as shown in Fig. 10. The detailed procedure is described as below:

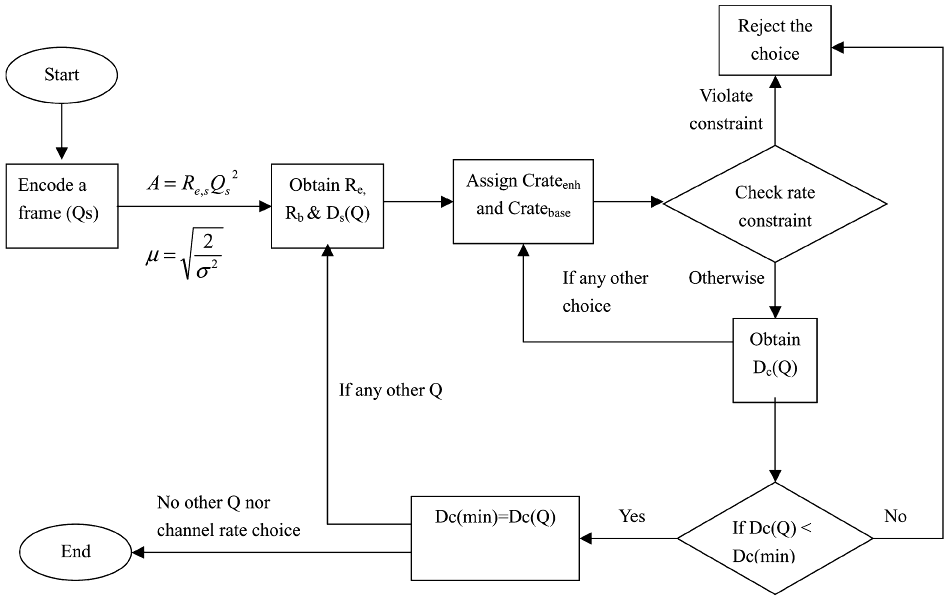

Fig. 10. Flow chart of the proposed adaptation algorithm.

(1) Encode the current frame by a sample quantization parameter Q_s. Count the bit rates generated from the base packets ($R_{\text{base,s}}$) and the enhancement packets ($R_{\text{enh,s}}$). Determine coefficient A via $A = R_{\text{enh,s}} \times Q_s^2$.
(2) For each enhancement packet, calculate the variance σ^2 of the residual DCT coefficients and the corresponding Laplacian parameter $\mu = \sqrt{\frac{2}{\sigma^2}}$.
(3) For each combination of (Q, C_{base}, C_{enh}), compute the total bit rate via Eq. (11). If it exceeds the given bit budget, reject the choice. If it is within the bit budget, compute the overall distortion via Eq. (10).
(4) Repeat Step 3 for all valid combinations and find out the best combination $(Q^*, C^*_{\text{base}}, C^*_{\text{enh}})$ that minimizes D.
(5) Quantize the DCT coefficients by Q^* and apply RCPC with rates C^*_{base} and C^*_{enh} to the base and the enhancement packets, respectively.
(6) Repeat Steps 2–5 to process all the packets in the frame.

5. Simulation results

5.1. Channel rate allocation for pre-compressed video bitstreams

We investigate the performance of the channel rate allocation system in Section 3 in high bit error rate environments. A binary symmetric channel with bit error rate (BER) 0.01 or 0.05 is simulated. For channel coding, RCPC with four rates, which are 8/8, 8/16, 8/24, and 8/32, is employed. Also, the Viterbi decoder is assumed to be capable of tracing back 80 symbols.

We use three test sequences. They are the ‘Claire’ 1st–50th frames, the ‘Foreman’ 1st–50th frames and the ‘Foreman’ 100th–150th frames, which have slow, moderate, and fast motion characteristics, respectively. Figs. 11–13 show the PSNR performances of the proposed algorithm (UEP) and the equal error protection (EEP) scheme, which uses the same channel code rate (= 8/16) for all base and enhancement packets. For error recovery, ‘w/ER’ indicates that the error concealment method in Section 2.2 is used, while ‘w/o ER’ denotes the direct copying algorithm. Since the locations of errors affect the quality of the reconstructed video significantly, each curve is obtained by averaging PSNRs over 100 different error patterns. It is clear that the method ‘UEP w/ER’ provides a significant performance improvement as compared to the other three methods.

For the slow motion sequence, the gap between UEP and EEP is larger than that between ‘w/ER’ and ‘w/o ER’ as shown in Fig. 11. This indicates that UEP is more powerful than the motion-compensated error concealment, since the simple copying

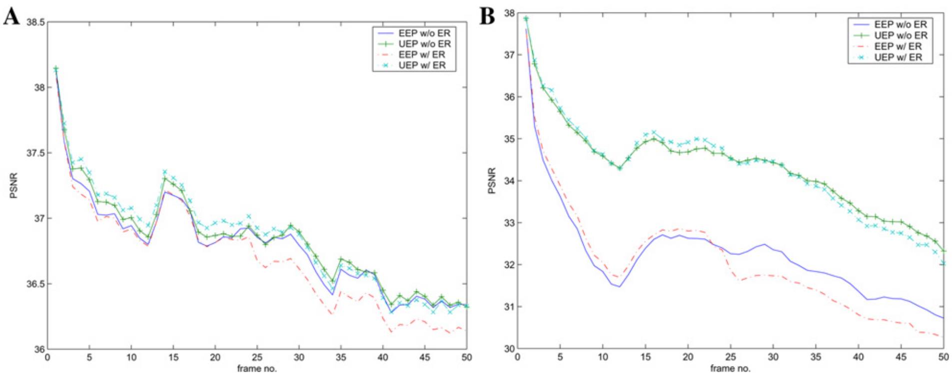

Fig. 11. PSNR comparison for the ‘Claire’ 1st–50th frames, which have slow motion characteristics: (A) BER = 0.01 and (B) BER = 0.05.

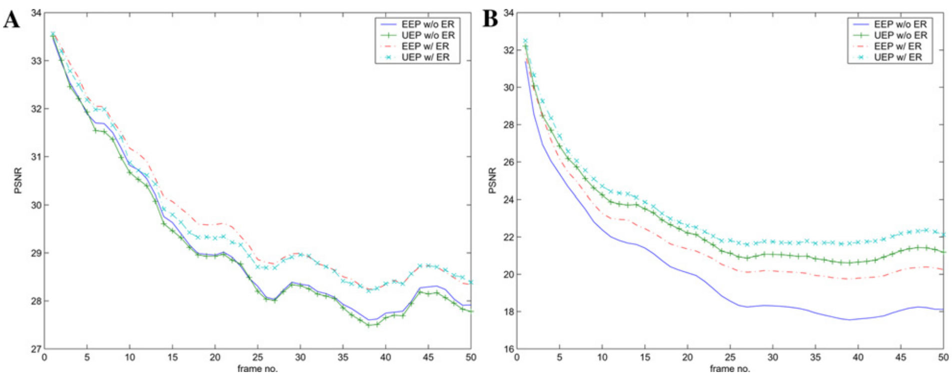

Fig. 12. PSNR comparison for the ‘Foreman’ 100th–150th frames, which have fast motion characteristics: (A) BER = 0.01 and (B) BER = 0.05.

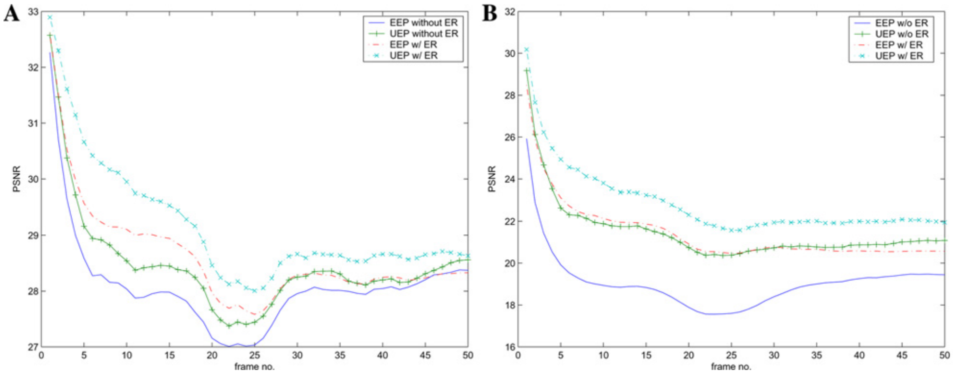

Fig. 13. PSNR comparison for the 'Foreman' 1st–50th frames, which have moderate motion activities: (A) BER = 0.01 and (B) BER = 0.05.

algorithm is sufficient to conceal the loss of slowly moving objects. However, even in the slow motion sequence, the importance of each packet varies greatly and UEP enhances the performance. On the contrary, for the fast motion sequence, the error concealment is more powerful than UEP when BER = 0.01, as shown in Fig. 12A. The simple copying algorithm causes severe discontinuities and artifacts if the sequence contains fast motion. Thus, the proposed motion-compensated concealment provides a much better performance than the simple copying algorithm. However, as BER increases, there are more packet losses and the error concealment performance becomes poorer. Thus, when BER = 0.05, UEP plays a more important role than the error concealment.

The best performance improvement is obtained in the moderate motion sequence as shown in Fig. 13. Both the error concealment and UEP provide significant PSNR improvements. As shown in Fig. 13B, UEP gives about 3 dB improvement, and the error concealment also gives about 3 dB improvement on the average. However, the total improvement is not additive. It is about 4 dB. That is because the error concealment tends to reduce the cost of the base packet loss. In other words, a good error concealment decreases the cost gap between the base packet and the enhancement packet, which in turn decreases the gain of UEP over EEP.

Fig. 14 compares several frames of the moderate motion sequence, which are reconstructed by the 'UEP w/ER' and 'EEP w/o ER' methods. It can be observed that the 'UEP w/ER' method provides much better image quality than 'EEP w/o ER.'

5.2. Adaptive source and channel rate allocation for real time video communications

Next, we investigate the performance of the proposed adaptive video transmission system in Section 4. In the following experiments, a binary symmetric channel with average BER = 0.004 and 0.02 is simulated. The spontaneous BER for a packet varies from 0.1 to 10^{-11} for an average BER of 0.004, and varies from 0.1 to 0.001 for an average BER of 0.02. RCPC with four channel code rates, 8/8, 8/16, 8/24, and 8/32, is employed and the Viterbi decoder traces back 80 symbols. The test

Fig. 14. Several frames of 'Foreman' sequence, which are reconstructed by (A) equal error protection and direct copying ('EEP w/o ER') and (B) unequal error protection and error concealment ('UEP w/ER').

video sequences are encoded in *IPPP*... format. The first I frame is assumed to be transmitted without any channel error, and the proposed algorithm is applied only to the following P frames.

Figs. 15–17 show the performances of the proposed adaptive system on 'Claire,' 'Salesman,' and 'Foreman' sequences, which are examples of slow, moderate, and fast motion pictures, respectively. For comparison, we also show the performance of the non-adaptive system under the same channel condition, where Q and the channel code rate are fixed to 20 and 8/16, respectively. Due to the bit rate control and the proper channel code rate assignment, the proposed algorithm reduces the packet loss rate significantly. It can be seen that the adaptive system provides at least 2 dB and up to 6 dB improvement as compared with the non-adaptive system.

For the slow motion sequence 'Claire,' the motion-compensated error concealment algorithm conceals corrupted regions faithfully. However, the non-adaptive system does not fully utilize the bandwidth, and its PSNR performance is limited by the quantization error. In contrast, the proposed algorithm assigns a smaller Q

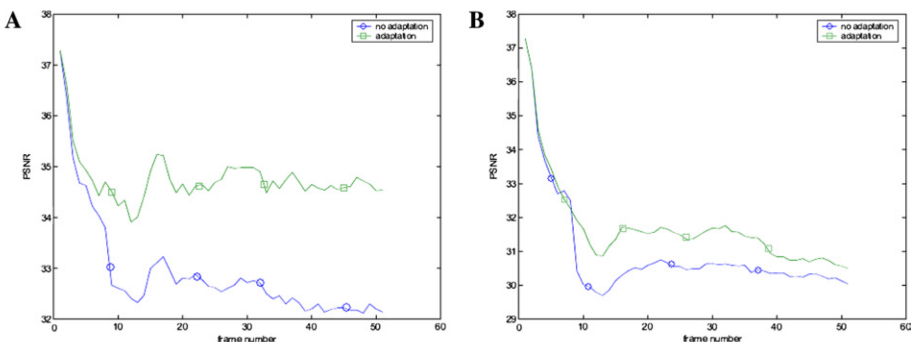

Fig. 15. The PSNR comparison for 'Claire' sequence at (A) BER $= 10^{-1}$ to 10^{-11} and (B) BER $= 10^{-1}$ to 10^{-3}.

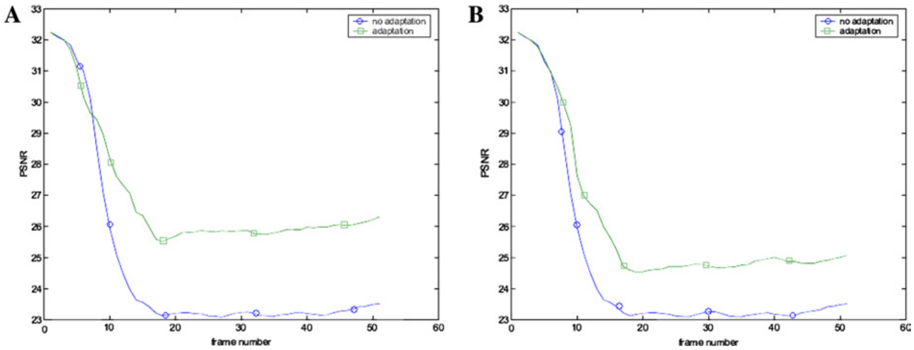

Fig. 16. The PSNR comparison for 'Salesman' sequence at (A) BER $= 10^{-1}$ to 10^{-11} and (B) BER $= 10^{-1}$ to 10^{-3}.

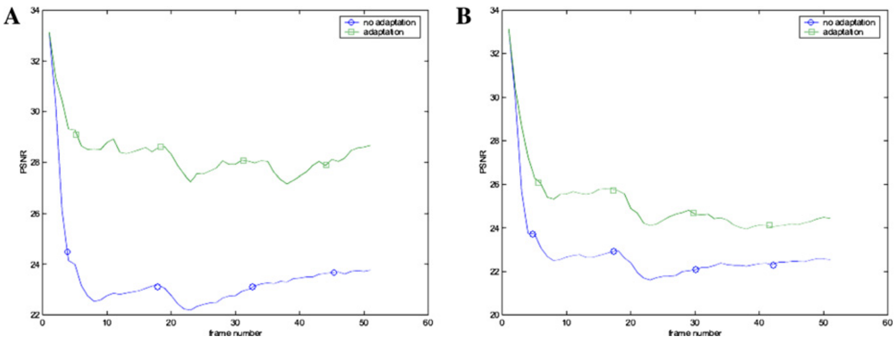

Fig. 17. The PSNR comparison for 'Foreman' sequence at (A) BER $= 10^{-1}$ to 10^{-11} and (B) BER $= 10^{-1}$ to 10^{-3}.

and improves the quality of the received video. For the moderate motion sequence 'Salesman,' the PSNR value decreases sharply around the 10th frame, where the sequence contains fast motion. Nevertheless, the proposed adaptive system still performs better than the non-adaptive system. For the fast motion sequence 'Foreman,' the average assigned Q value is around 20. Thus, the average source bit rate in the adaptive system is close to that in the non-adaptive system. However, the adaptive system allocates different channel rates to the base packets and the enhancement packets to minimize the overall distortion. Due to the fast motion activities of 'Foreman' sequence, the base packets are more important, thus being protected with stronger channel codes. Therefore, the adaptive system provides much better PSNR performance than the non-adaptive system as shown in Fig. 17A. As shown in Fig. 17B, when the channel condition becomes worse, the adaptive system has to lower Q and the channel code rate to satisfy the overall bit rate constraint. Therefore, the PSNR improvement becomes smaller.

Fig. 18. Several frames of 'Foreman' sequence, which are reconstructed by (A) the non-adaptive system and (B) the proposed adaptive system.

Fig. 18 shows examples of reconstructed frames of 'Foreman' sequence when BER varies from 10^{-1} to 10^{-11}. In the non-adaptive system, the reconstructed frames contain severe distortions and blurring. On the contrary, the proposed algorithm reconstructs the frames with an acceptable image quality. These simulation results indicate that the proposed algorithm is an effective method for robust video transmission.

6. Conclusions

In this work, we developed a video coder based on layered coding and interleaved packetization. In addition, we proposed joint source/channel coding schemes for two typical video communication scenarios. The first one was designed for pre-compressed video bitstreams, such that the expected mean square error is minimized subject to a constraint on the overall bit rate. The second scheme was designed for real-time video transmission over wireless channels, where the source and the channel rates of each packet are jointly optimized to maximize the quality of reconstructed video. Simulation results demonstrated that the proposed algorithms for both scenarios provide acceptable image quality even in high bit error rate environments.

Appendix A. Derivation of quantization distortion

When a uniform quantizer is employed, the expected distortion is expressed as

$$D_{\text{source}} = \sum_{i=-\infty}^{\infty} \int_{q(i-1/2)}^{q(i+1/2)} (x - qi)^2 p(x)\,\mathrm{d}x,$$

where q denotes the quantization step size and i denotes the quantization index. Suppose that x has the Laplacian distribution in Eq. (5). In real implementation, i can not be infinitely large. We assume that i is restricted between -127 and 127. Then, D_{source} is derived as

$$D_{\text{source}} = 2\int_0^{1/2} x^2 \frac{\mu}{2} \mathrm{e}^{-\mu x}\,\mathrm{d}x + 2\sum_{i=1}^{127} \int_{q(i-1/2)}^{q(i+1/2)} (x-qi)^2 \frac{\mu}{2} \mathrm{e}^{-\mu x}\,\mathrm{d}x$$
$$+ 2\int_{127.5q}^{\infty} (x-128q)^2 \frac{\mu}{2} \mathrm{e}^{-\mu x}\,\mathrm{d}x = \frac{2-\mathrm{e}^{-1/2\mu q}(2+\mu q+\frac{1}{4}\mu^2 q^2)}{\mu^2}$$
$$+ \sum_{i=1}^{127} \frac{\mathrm{e}^{-\mu q(i+1/2)}(-2-\mu q-\frac{1}{4}\mu^2 q^2) - \mathrm{e}^{-\mu q(i-1/2)}(-2+\mu q-\frac{1}{4}\mu^2 q^2)}{\mu^2}$$
$$+ \frac{\mathrm{e}^{-127.5\mu q}(2-\mu q+\frac{1}{4}\mu^2 q^2)}{\mu^2} = \frac{2}{\mu^2}\left[1 - \frac{\mathrm{e}^{-1/2\mu q} - \mathrm{e}^{-128.5\mu q}}{1-\mathrm{e}^{-\mu q}}(\mu q)\right].$$

References

[1] Information Technology—Generic Coding of Audio-Visual Objects, Part 2: Visual, ISO/IEC 14496-2, 1998.

[2] Video Coding for Low Bitrate Communication, ITU-T Recommendation H.263, 1998.

[3] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley, New York, 1991.

[4] M. Gallant, F. Kossentini, Rate-distortion optimized layered coding with unequal error protection for robust internet video, IEEE Trans. Circuits Syst. Video Technol. 11 (2001) 357–372.

[5] H. Li, C.W. Chen, Robust image transmission with bidirectional synchronization and hierarchical error correction, IEEE Trans. Circuits Syst. Video Technol. 11 (2001) 1183–1187.

[6] H. Zheng, K.R. Liu, Robust image and video transmission over spectrally shaped channels using multicarrier modulation, IEEE Trans. Multimedia 1 (1999) 88–103.

[7] H. Gharavi, S.M. Alamouti, Multipriority video transmission for third-generation wireless communication systems, Proc. IEEE 87 (1999) 1751–1763.

[8] W.R. Heinzelman, M. Budagavi, R. Talluri, Unequal error protection of MPEG-4 compressed video, in: Proc. ICIP, vol. 2, 1999, pp. 530–534.

[9] J. Hagenauer, Rate-compatible punctured convolutional codes (RCPC codes) and their applications, IEEE Trans. Commun. 36 (1988) 389–400.

[10] L.-F. Wei, Coded modulation with unequal error protection, IEEE Trans. Commun. 41 (1993) 1439–1449.

[11] H. Gharavi, Pilot-assisted 16-level QAM for wireless video, IEEE Trans. Circuits Syst. Video Technol. 12 (2002) 77–89.

[12] G. Cheung, A. Zakhor, Bit allocation for joint source/channel coding of scalable video, IEEE Trans. Image Process. 9 (3) (2000) 340–356.

[13] L.P. Kondi, A.K. Katsagglos, Joint source-channel coding for scalable video using models of rate-distortion functions, in: Proc. ICASSP, vol. 3, 2001, pp. 1377–1380.

[14] B. Hochwald, K. Zeger, Tradeoff between source and channel coding, IEEE Trans. Inform. Theory 43 (5) (1997) 1412–1424.

[15] L.P. Kondi, S.N. Batalama, D.A. Pados, A.K. Katsagglos, Joint source-channel coding for scalable video over DS-CDMA multipath fading channels, in: Proc. ICIP, vol. 1, 2001, pp. 994–997.

[16] C.-Y. Hsu, A. Ortega, M. Khansari, Rate control for robust video transmission over burst-error wireless channels, IEEE J. Select. Areas Commun. 17 (1999) 1–18.

[17] Q. Zhang, W. Zhu, Z. Ji, Y.-Q. Zhang, A power-optimized joint source channel coding for scalable video streaming over wireless channel, in: Proc. ISCAS, vol. 5, 2001, pp. 137–140.

[18] Z. Ji, Q. Zhang, W. Zhu, Y.-Q. Zhang, End-to-end power-optimized video communication over wireless channels, in: Proceedings of the IEEE Workshop on Multimedia Signal Processing, 2001, pp. 447–452.
[19] A. Ortega, K. Ramchandran, Rate-distortion methods for image and video compression, IEEE Signal Process. Magazine 15 (1998) 23–50.
[20] K. Stuhlmuller, N. Farber, M. Link, B. Girod, Analysis of video transmission over lossy channels, IEEE J. Select. Areas Commun. 18 (2000) 1012–1032.
[21] K.H. Yang, A. Jacquin, N.S. Jayant, Rate-distortion optimized video coding considering frameskip, in: Proc. ICIP, 2001, pp. 534–537.
[22] K.H. Yang, A. Jacquin, N.S. Jayant, A normalized rate-distortion model for H.263-compatible codecs and its application to quantizer selection, in: Proc. ICIP, 1997, pp. 41–44.
[23] Z. He, S.K. Mitra, Novel rate-distortion analysis framework for bit rate and picture quality control in DCT visual coding, IEE Proc. Vision, Image, Signal Process. (2001) 398–406.
[24] T. Chiang, Y.-Q. Zhang, A new rate control scheme using quadratic rate distortion model, IEEE Trans. Circuits Syst. Video Technol. 7 (1997) 246–250.
[25] Z. He, Y.K. Kim, S. Mitra, Low-delay rate control for DCT video coding via ρ-domain source modeling, IEEE Trans. Circuits Syst. Video Technol. 11 (2001) 928–940.
[26] M. Bystrom, T. Stockhammer, Modeling of operational distortion-rate characteristics for joint source-channel coding of video, in: Proc. ICIP, 2000, pp. 359–362.
[27] H.-M. Hang, J.-J. Chen, Source model for transform video coder and its application-part I: fundamental theory, IEEE Trans. Circuits Syst. Video Technol. 7 (1997) 287–298.
[28] J. Ribas-Corbera, S. Lei, Rate control in DCT video coding for low-delay communications, IEEE Trans. Circuits Syst. Video Technol. 9 (1999) 172–185.
[29] E.Y. Lam, J.W. Goodman, A mathematical analysis of the DCT coefficient distributions for images, IEEE Trans. Image Process. 9 (2000) 1661–1666.