

Analytic End-to-End Rate Distortion Modeling and Control for Packet Video Over Wireless Network

Zhihai He, Jianfei Cai, and Chang Wen Chen
Interactive Media Group
Sarnoff Corporation, Princeton NJ 08543
{zhe, cchen}@sarnoff.com

March 15, 2002

Abstract

In this work, we analyze the effect of packet loss caused by channel errors on the decoded video quality at the receiver end, and propose an analytic channel distortion model. Coupled with the source coding rate-distortion (R-D) model developed in our previous work [1], a low-complexity end-to-end R-D model is developed for packet video coding and transmission over wireless network. This end-to-end R-D model enables us to develop an optimal joint source-channel bit allocation and rate control algorithm. The proposed algorithm is able to choose appropriate coding parameters for the source and channel encoders, which results in significantly improved video presentation quality at the receiver end.

I. Introduction

With the increasing bandwidth in the third generation (3G) mobile networks and growing demand for visual communication, packet video transmission over wireless network becomes possible and has received much attention during the past few years. Due to the limited bandwidth of the wireless channels, video signals need to be highly compressed using some very efficient video coding algorithms, such as H.263 [2] and MPEG-4 [3]. On the other hand, due to the error-prone nature of wireless channels, error control techniques such as forward error correction (FEC) and automatic repeat request (ARQ) are desired to add controlled redundancy to ensure reliable video transmission. Because of stringent delay constraint for real-time video transmission, it is often considered more beneficial to apply FEC than to use ARQ.

In a FEC-based video transmission system, one of the key problems is joint source-channel rate control. Given channel conditions, such as transmission bandwidth and channel bit error rate (BER), the encoder needs to find out the optimal bandwidth allocation between source coding and channel coding so that the overall video quality degradation at the receiver end is minimized. Note that an end-to-end video coding and transmission system consists of two major components: source coding and network transmission. Correspondingly, the picture distortion at the receiver

end is mainly caused by quantization errors in source coding and channel errors (or packet loss) in network transmission. These two types of distortion are called “source distortion” and “channel distortion”, respectively. The source distortion analysis involves the modeling of the complex encoding process. In channel distortion analysis, we need to model the video transmission related operations, such as bits stream packetization, transmission errors, channel coding and decoding, video decoding, error concealment, etc.

A. Related Works

For a given source coding bit rate R_s , to estimate the corresponding source distortion D_s , we need to model and analyze the R-D behavior of the video encoder. Due to the large variation in the characteristics of input videos and sophisticated data representation schemes employed by the coding algorithm, accurate modeling and control of the R-D behavior of the video encoder remains a challenging problem. In the literature, operational R-D estimation is often used, where the R-D curve is constructed from actual R-D measurements [4, 5]. To reduce the computational complexity, in MPEG-4 VM7 [6] and H.263 TMN8 [7] rate control algorithms, the coding statistics of previous frames or macroblocks (MBs) are employed to estimate the R-D model parameters for the current frame or MB.

Standard video coding, such as H.263 and MPEG-4, employ a motion compensation based discrete cosine transform (MC-DCT) coding scheme. While motion compensation significantly improves the coding efficiency, it also causes error propagation when transmission errors occur, which may significantly degrade the picture quality at the receiver end. In channel distortion analysis, we need to model this inter-frame error propagation. In addition, the channel distortion analysis also needs to consider the specific source/channel decoding schemes, packetization method, patterns of the channel errors, error concealment, etc. Several approaches for channel distortion estimation have been proposed in the literature [8, 9]. To analyze the video transmission over lossy channel, a heuristic approach is introduced in [8], where the channel distortion formula is derived based on a leaking filtering model of the video decoder. An experimental approach using statistical simulation of the video decoder at the encoder side is employed in [9] to estimate the channel distortion under error concealment. Such scheme involves potentially high computational complexity and implementation cost, which is prohibited in wireless video communication.

B. Proposed Work

In this paper, based on the statistical analysis of the inter-frame error propagation, error concealment, and channel decoding, we develop an analytic formula for the channel distortion. Coupled with the accurate and robust source R-D models developed in our previous work [1], a joint source-channel rate control scheme is proposed for wireless video coding and transmission. This end-to-end R-D analysis framework can be applied to any standard DCT-based motion-compensated video coding scheme and any video sequence. The simulations show that the optimal joint source-channel rate control can achieve up to 1.5 dB PSNR gain, comparing with the conventional threshold-based scheme.

The rest of the paper is organized as follows. Section II reviews the source R-D models and rate control algorithm developed in our previous work [1]. The channel distortion model and the adaptive estimation scheme are presented in Section III. In Section IV, based on the source and channel R-D models, we propose an optimal joint source-channel bit rate allocation and control algorithm. The experimental results are presented in Section V. Some concluding remarks are given in Section VI.

II. R-D Analysis of Source Coding

In our previous work [1], we have developed a robust and accurate R-D model for DCT-based video coding. Specifically, in this model, we consider the source coding bit rate R_s and distortion D_s as functions of ρ defined to be the percentage of zeros among the quantized DCT coefficients. This consideration is based on the following observation. In the classical R-D analysis, R_s and D_s are treated as functions of the quantization parameter (or step size) q . We notice that in standard DCT-based video coding ρ monotonically increases with q . (In other words, the increase of q will typically result in more zeros among the quantized DCT coefficients.) This implies that there is a one-to-one mapping between ρ and q . Therefore, mathematically, R_s and D_s are also function of ρ . We observe that, in the ρ domain, the R-D functions have unique behavior. Specifically, R_s has a linear relationship with ρ . Mathematically, we have

$$R_s(\rho) = \theta \cdot (1 - \rho) + C_h, \quad (1)$$

where θ is a constant, and C_h represents the number of bits for header information and motion vectors which do not depend on the quantization step size. In the ρ -domain, we have also developed the following source distortion model,

$$D_s(\rho) = \sigma^2 e^{-\alpha(1-\rho)}, \quad (2)$$

where σ^2 is the variance of the source data and α is a constant. Our extensive experimental results in [1] have shown that the above R-D models are very accurate. Based on the rate model (1), a linear rate control algorithm has also been developed, with which we can control the video encoder to achieve the target bit rate accurately and robustly. A detailed treatment of these R-D models and corresponding rate control algorithm can be found in [1].

III. Channel Distortion Analysis

In wireless video transmission, channel coding is used to correct bit errors in the compressed video data stream. Due to channel error mismatch and the limited error correction capacity of the channel decoder, bits errors often still exist even after error correction. When a corrupted codeword cannot be correctly decoded, the video decoder will jump to the next slice starting with a re-synchronization mark with all the intermediate bits being skipped. In other words, the whole packet is lost. We denote the packet loss ratio as p . If we assume each packet contains the same number of macroblocks or pixels, then the loss ratio of pixels is also p .

Let $F(n, i)$ be the original value of pixel i in the n -th video frame, and $\hat{F}(n, i)$ be the corresponding reconstruction value in the feedback loop at the encoder. We denote the reconstruction value at the receiver end as $\tilde{F}(n, i)$. For intercoded macroblocks (MBs), let $e(n, i)$ be the motion compensation difference at the encoder. Let $\hat{e}(n, i)$ and $\tilde{e}(n, i)$ be the corresponding reconstruction values at the encoder and decoder, respectively. Due to the randomness of channel errors, $\tilde{F}(n, i)$ and $\tilde{e}(n, i)$ are random variables. Therefore, we can only model and analyze the expected picture distortion at the receiver end which is given by

$$D(n) = E \left\{ [F(n, i) - \tilde{F}(n, i)]^2 \right\}. \quad (3)$$

Meanwhile, we define source coding distortion $D_s(n)$ and channel distortion $D_c(n)$ as

$$D_s(n) = E \left\{ [F(n, i) - \hat{F}(n, i)]^2 \right\}, \quad (4)$$

$$D_c(n) = E \left\{ [\hat{F}(n, i) - \tilde{F}(n, i)]^2 \right\}. \quad (5)$$

In the following experiment, we show the quantization error and the channel error are uncorrelated. We code the “Foreman” QCIF video at 96 kbps and 15 fps with MPEG-4 and simulate the transmission with packet loss at a loss rate of 2%. In Fig. 1, we plot the $D(n)$ and $D_s(n) + D_c(n)$ for each frame. It can be seen that $D(n)$ is approximately equal to $D_s(n) + D_c(n)$. This implies the quantization error and channel error are uncorrelated with each other. Using the R-D models presented in Section II, we can accurately estimate $D_s(n)$. Therefore, the only thing left is to estimate $D_c(n)$.

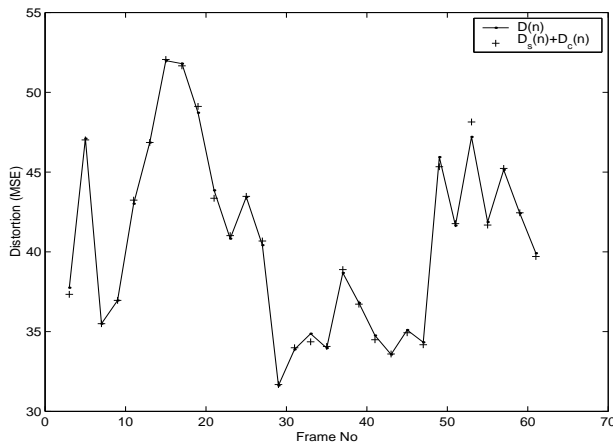


Figure 1: Comparison between the overall distortion $D(n)$ and $D_s(n) + D_c(n)$ for “Foreman” QCIF video coded at 96 kbps and 15 fps and a loss rate of 2%.

A. Channel Distortion Model

At the decoder side, we assume the following error concealment scheme: If a MB is skipped by the decoder, both the motion vectors and the texture information are lost. The decoder simply copies the MB at the same location from the previous decoded frame. Based on this error concealment

scheme, we provide a statistical analysis of the channel distortion. For pixels in intracoded MBs, the expected channel distortion is given by

$$\begin{aligned}
D_c^I(n) &= E \left\{ [\hat{F}(n, i) - \tilde{F}(n, i)]^2 \right\} \\
&= (1 - p) \cdot E \left\{ [\hat{F}(n, i) - \hat{F}(n, i)]^2 \right\} + p \cdot E \left\{ [\hat{F}(n, i) - \tilde{F}(n - 1, i)]^2 \right\} \\
&= p \cdot E \left\{ [\hat{F}(n, i) - \hat{F}(n - 1, i)]^2 \right\} + p \cdot E \left\{ [\hat{F}(n - 1, i) - \tilde{F}(n - 1, i)]^2 \right\} \\
&= p \cdot RFD(n, n - 1) + p \cdot D_c(n - 1),
\end{aligned} \tag{6}$$

where $RFD(n, n - 1)$ represents the mean square error (MSE) between the reconstructed frames n and $n - 1$. It should be noted that the third identity in (6) is based on uncorrelation between the frame difference and the channel distortion. Note that, before quantization and coding, $\hat{F}(n, i)$ is not available. Therefore, $RFD(n, n - 1)$ can not be evaluated directly. However, we do know the MSE between the original frames n and $n - 1$, defined as

$$\mathcal{F}_d(n, n - 1) = E \left\{ [F(n, i) - F(n - 1, i)]^2 \right\}. \tag{7}$$

If we assume

$$RFD(n, n - 1) = a \cdot \mathcal{F}_d(n, n - 1), \tag{8}$$

where a is a constant, (6) then becomes

$$D_c^I(n) = ap \cdot \mathcal{F}_d(n, n - 1) + p \cdot D_c(n - 1). \tag{9}$$

Heuristically, the encoder can be regarded as a low-pass filter [8] which removes the information and reduces the overall energy in the original picture. From (8) we can see that the constant a describes the energy reduction of this low-pass filter.

For a pixel in intercoded MBs, in case of no channel errors, its reconstruction value is $\hat{e}(n, i) + \tilde{F}(n - 1, j)$ where pixel j is the motion prediction of pixel i . If the MB is lost, the reconstruction value of pixel i is $\tilde{F}(n - 1, i)$, which is copied from the previous decoded frame. Therefore, we have

$$\begin{aligned}
D_c^P(n) &= E \left\{ [\hat{F}(n, i) - \tilde{F}(n, i)]^2 \right\} \\
&= (1 - p) \cdot E \left\{ [\hat{F}(n, i) - \hat{e}(n, i) - \tilde{F}(n - 1, j)]^2 \right\} + p \cdot E \left\{ [\hat{F}(n, i) - \tilde{F}(n - 1, i)]^2 \right\} \\
&= (1 - p) \cdot E \left\{ [\hat{F}(n - 1, j) - \tilde{F}(n - 1, j)]^2 \right\} + p \cdot RFD(n, n - 1) + p \cdot D_c(n - 1).
\end{aligned} \tag{10}$$

Note that $\{\hat{F}(n - 1, j)\}$ is the motion prediction frame. Obviously, it is different from $\{\hat{F}(n - 1, i)\}$ because different pixels in frame n may refer to the same pixels in frame $n - 1$. If we assume

$$E \left\{ [\hat{F}(n - 1, j) - \tilde{F}(n - 1, j)]^2 \right\} = b \cdot D_n(n - 1) \tag{11}$$

where b is a constant, we have

$$D_c^P(n) = [(1 - p)b + p] \cdot D_c(n - 1) + pa \cdot \mathcal{F}_d(n, n - 1). \tag{12}$$

In frame n , let M be the total number of MBs and L be the number of intracoded MBs. $\beta = \frac{L}{M}$ is then the intra refreshing rate. The overall channel distortion can be approximated by

$$\begin{aligned} D_c(n) &= \beta D_c^I(n) + (1 - \beta) D_c^P(n) \\ &= [(1 - \beta)(1 - p)b + p] \cdot D_c(n - 1) + pa \cdot \mathcal{F}_d(n, n - 1) \\ &= \Gamma_1 \cdot D_c(n - 1) + \Gamma_2 \cdot \mathcal{F}_d(n, n - 1), \end{aligned} \quad (13)$$

where

$$\Gamma_1 = (1 - p)b + p, \quad \Gamma_2 = pa. \quad (14)$$

Solving the recursive equation (13), we have,

$$D_c(n) = \Gamma_1^n \cdot D_c(0) + b \sum_{k=1}^n \Gamma_1^k \cdot \mathcal{F}_d(k, k - 1). \quad (15)$$

In wireless video communication over noisy channels, if the feedback information on the channel condition and transmission status is given, the encoder knows the decoded picture quality of frame $n - d$ and its previous frames, where Δ is the feedback delay (in the unit of frame interval). In other words, $\{D_c(n - \Delta - m) | m \geq 0\}$ are available at the encoder through the channel feedback information. Eq. (13) or (15) is then used to calculate $D_c(n)$. The model parameters a and b can also be adaptively estimated or adjusted with feedback information.

B. Estimate Pixel Loss Rate

In wireless video communication over a noisy channel, the channel introduces bit errors into the coded video bits stream. In this paper, we consider a random binary symmetric channel (BSC) model and use (N, K) Reed-Solomon (RS) block codes with 8 bits per symbol. The code rate $r = K/N$ is determined by the joint source-channel bit allocation scheme proposed in Section IV. Note that in the RS codes, any error pattern with no more than

$$T = \lfloor \frac{N - K}{2} \rfloor \quad (16)$$

symbol errors can be corrected. We denote the symbol error rate (SER) as \mathcal{E} . The decoded symbol error rate is then given by

$$\mathcal{E}_d = 1 - \sum_{i=0}^K \sum_{j=0}^{N-K} \binom{K}{i} \binom{N-K}{j} \mathcal{E}^{i+j} (1 - \mathcal{E})^{N-i-j} \cdot \eta(i, j), \quad (17)$$

where

$$\eta(i, j) = \begin{cases} 1, & \text{if } i + j \leq T; \\ (K - i)/K, & \text{otherwise.} \end{cases} \quad (18)$$

Once a symbol cannot be corrected by the RS decoder, it will be detected by the video decoder due to the syntax violations. In this case, the decoder will jump to the next slice starting with a re-synchronization mark and skips all the intermediate symbols. Suppose the slice has L symbols. The loss rate of the slice is determined by

$$p = 1 - (1 - \mathcal{E}_d)^L, \quad (19)$$

which is used by the channel distortion model (13).

IV. Joint Source-Channel Bit Allocation

To minimize the overall picture distortion at the receiver end, we need to adjust the parameters of the video encoder and the RS codes for different input video sequences and different channel conditions. Specifically, given a channel bandwidth R_T , we need to perform optimal bit allocation between source and channel coding. From (13) we can see that the channel distortion D_c is a function of p , which is in turn a function of the channel coding rate $R_c = (1 - r)R_T$. Therefore, the optimal joint source-channel bit allocation can be formulated as follows,

$$\min_{0 \leq r \leq 1} D(r) = D_c[(1 - r) \cdot R_T] + D_s[r \cdot R_T]. \quad (20)$$

Based on the R-D estimation schemes presented in Section 2 and 3, an optimal bit allocation algorithm is proposed as follows:

- **Step 1.** *Estimate source coding distortion.* With (1) and (2), estimate the source coding D-R function $D_s(R_s)$ which is stored as a look-up table.
- **Step 2.** *Estimate channel distortion.* Calculate the frame difference $\mathcal{F}_d(n, n - 1)$. Note that the frame differences of all the previous frames are already available. Based on the previous picture quality information $\{D_c(n - d - m), p\}$, using Eqs. (13) and (19), estimate the decoded picture distortion of the current frame n for any given RS coding rate r .
- **Step 3.** *Determine the optimal coding rate.* Find r which minimize $D(r)$ in (20).
- **Step 4.** *Encoder rate control.* The target bit rate for source coding is given by $r \cdot R_T$. The linear rate control algorithm proposed in [1] is employed to achieve this target bit rate.
- **step 5.** *Update model parameters.* Based on the source coding statistics, update the model parameters θ and α as described in [1]. Based on the new feedback information from the decoder, update the channel distortion model parameters a and b .

V. Experimental Results

We have implemented the above R-D estimation and optimum bit allocation algorithm in the MPEG-4 codec. In our simulations, each frame is partitioned into 5 packets with approximately equal number of MBs. With MB interleaving, each packet contains every fifth MB in the raster scan order, and starts with a re-synchronization mark. Such type of MB interleaving helps the error concealment at the decoder side [10]. The coded video data stream is further coded by RS codes. A random bit error generator is used to produce bit errors at a specified error rate. At the receiver end, the corrupted bits stream is first RS decoded to correct the bit errors. The temporal replacement concealment scheme is applied at the video decoder. For different input video sequences and channel conditions, we test the proposed joint source-channel bit allocation and rate control algorithm, and compare it with the conventional threshold-based bit allocation scheme.

First, we test the accuracy of the channel distortion model (13). To this end, we run the modified MPEG-4 codec over “Carphone” QCIF video at 96 kbps without Intra-frame refreshing.

In other words, the channel distortion in the first P-frame propagates to the end of the sequence without being stopped. In Fig. 2, we plot actual channel distortion and the estimated one for each frame at various feedback delays (1, 3, 5, 9, 15 and 20 frames). It can be seen that the estimation is very accurate even if we use the feedback information of many frame intervals ago.

Next, we demonstrate the performance of the proposed adaptive rate allocation and control algorithm. In Fig. 3, we plot the decoded picture quality for “Foreman” QCIF video coded at 96 kbps and 15 bps when the proposed adaptive rate control algorithm and the conventional threshold-based bit allocation algorithm are applied. For the threshold-based scheme, the SER threshold is set to 0.1%. In Table I, we summarize the PSNR results for different input video sequence and different SER threshold settings. Fig. 4 show the Frame 100 in decoded videos when the proposed algorithm (right) and the conventional threshold-based scheme (left) are applied. It can be seen that with adaptive rate allocation and control, the video encoder always chooses appropriate source/channel coding bit rates and encoder settings, which yields significantly improved picture quality at the receiver end.

VI. Conclusion

In this paper, we have presented a statistical analysis of the picture distortion introduced by channel errors, namely, channel distortion. Compared to other channel distortion estimation models, this scheme has much lower computational complexity and implementation cost, which is highly desirable in wireless video communication applications. Coupled with the source coding rate-distortion (R-D) models developed in [1], an optimal joint source-channel bit allocation and rate control algorithm is proposed. Our experimental results show that the proposed adaptive rate allocation and control algorithm significantly improves the decoded picture quality.

References

- [1] Z. He, Y. Kim, and S. K. Mitra, “Object-level bit allocation and scalable rate control for MPEG-4 video coding,” *Proceedings of Workshop and Exhibition on MPEG-4*, San Jose, CA, USA, June 2001.
- [2] ITU-T, “Video coding for low bit rate communications,” *ITU-T Recommendation H.263*, version 1, version 2, January 1998.
- [3] MoMuSys codec, “MPEG4 verification model version 7.0,” *ISO / IEC JTC1 / SC29 / WG11 Coding of Moving Pictures and Associated Audio MPEG97*, Bristol, U.K., March 1997.
- [4] W. Ding and B. Liu, “Rate control of mpeg video coding and recording by rate-quantization modeling,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, pp. 12–20, February 1996.

Table 1: Picture quality comparison for different video sequences and SER thresholds (with channel SER 0.1%)

Video Name	Channel Bandwidth	PSNR This work	PSNR Threshold-Base		
			0.05%	0.1%	0.2%
Foreman	96	31.40	30.23	30.71	29.81
Flowergarden	256	30.20	29.02	29.26	27.60
Salesman	96	34.40	34.21	33.28	33.52
Akiyo	64	36.98	36.75	36.24	36.56

- [5] L.-J. Lin and A. Ortega, “Bit-rate control using piecewise approximated rate-distortion characteristics,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 38, pp. 82–93, January 1990.
- [6] T. Chiang, Y. -Q. Zhang, “A new rate control scheme using quadratic rate distortion model,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol.7, pp. 246 – 250, February 1997.
- [7] J. Ribas-Corbera and S. Lei, “Rate control in DCT video coding for low-delay communications,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9, pp. 172 – 185, February 1999.
- [8] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, “Analysis of video transmission over lossy channels,” *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1012 – 1032, June 2000.
- [9] R. Zhang, S. L. Regunathan, and K. Rose, “Video coding with optimal Inter/Intra-mode switching for packet loss resilience,” *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 966 – 976, June 2000.
- [10] Y. Wang, Q.-F. Zhu, and L. Shaw, Maximally Smooth Image Recovery in Transform Coding, *IEEE Trans. Commun.*, vol. 41, no. 10, Oct. 1993, pp. 1544–1551.

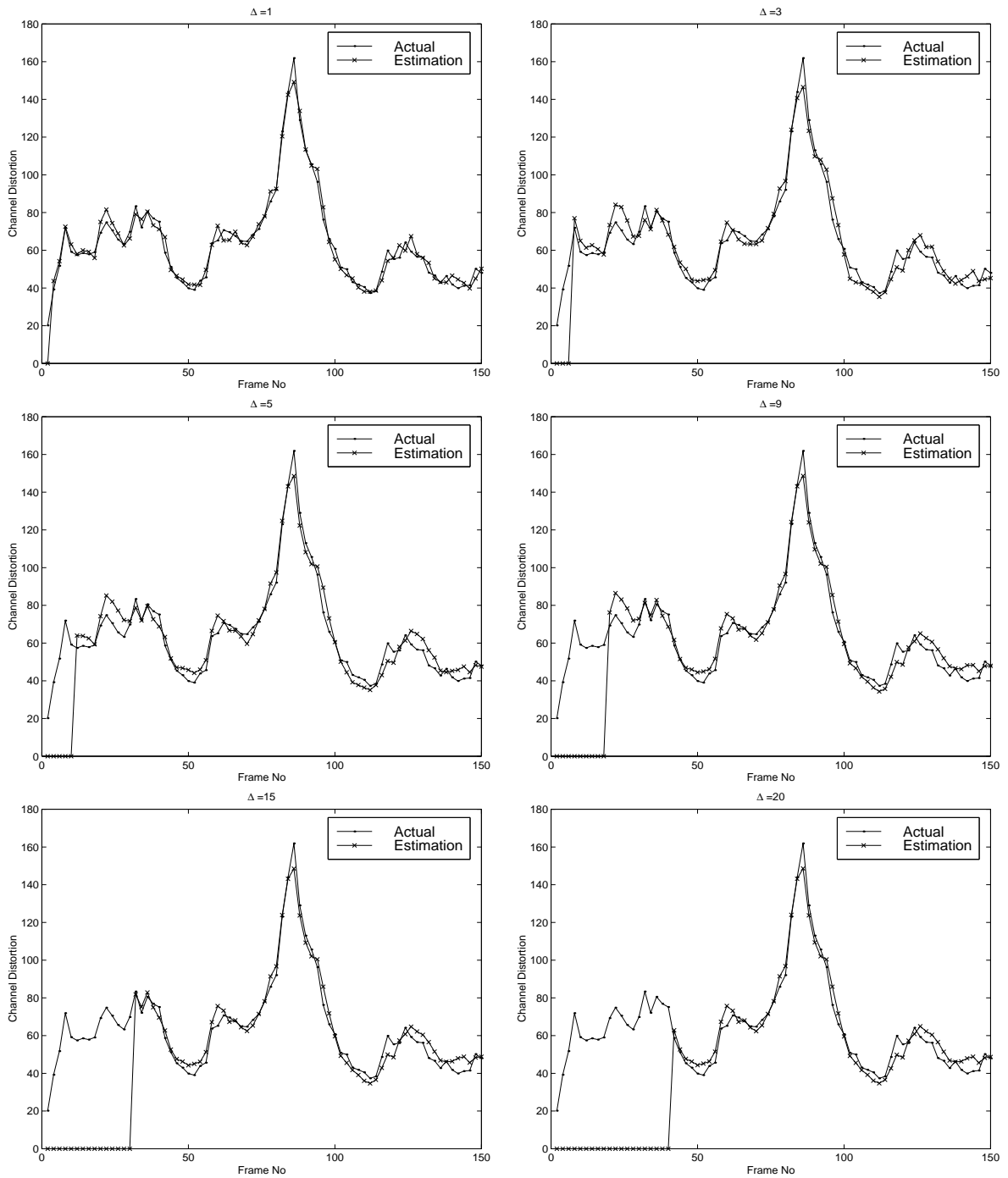


Figure 2: Channel distortion estimation results for “Carphone” QCIF video coded at 96 kbps, 15 fps and a packet loss ration of 5%. The title of each sub-figure shows the number of frames in feedback delay.

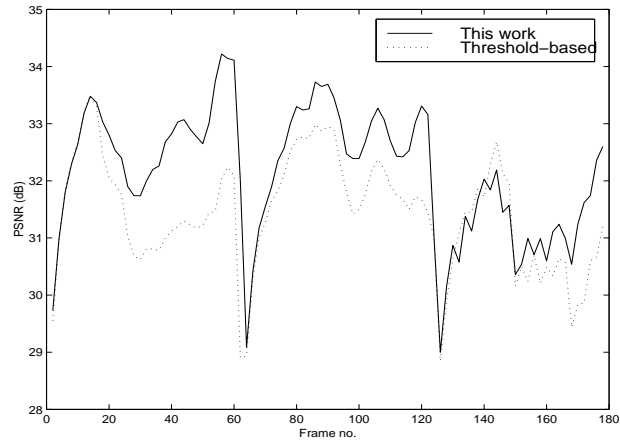


Figure 3: Decoded picture quality comparison when the proposed joint source-channel bit allocation scheme and the threshold-base scheme are applied. The video sequence is “Foreman” coded at 96 kbps, 15 bps and a BER of 0.01%



Figure 4: Comparison between the decoded pictures when the proposed joint source-channel bit allocation scheme (right) and the threshold-base scheme (left) are applied. The video sequence is “Carphone” coded at 64 kbps, 15 bps and a BER of 0.01%