OPTIMAL RATE ALLOCATION FOR VIDEO CODING AND TRANSMISSION OVER WIRELESS CHANNELS

Jie Song

Lucent Technologies
101, Crawfords Corner Road
Holmdel, NJ 07733

K. J. Ray Liu

Electrical & Computer Engineering Department
Institute for System Research
University of Maryland, College Park, MD 20742

ABSTRACT

An integrated framework of optimal rate allocation for video coding is presented in the case of transmission over wireless channels without feedback channels. For a fixed channel bit rate and finite number of channel coding rate, the proposed scheme can find the optimal source and channel coding pair and corresponding robust video coding scheme such that the expected end-to-end distortion of video signals can be minimized. With the assumption that encoder has the stochastic channel information, the proposed scheme takes into account robust video coding, channel coding and packetization, error concealment techniques altogether. Simulation results show the accuracy and optimality of the proposed method.

1. INTRODUCTION

Multimedia applications including video-phone, video conferencing will be provided in the third generation (3G) wireless systems. For real time applications, delay constraint makes the conventional Automatic Repeat reQuest (ARQ) and deep interleaver not suitable. Feedback channel can be used to deal with the error effects incurred in image and video transmission over error-prone channels, but in applications such as broadcasting services, there is no feedback channel available. In this case, the optimal tradeoff between source and channel coding rate allocation for video transmission over error-prone channels should be considered. Therefore, Joint Source and Channel Coding (JSCC) becomes an important research topic in the last decades [1, 2]. Most JSCC schemes focus on image and ideal signal models [1, 3, 2]. For video coding and transmission, many works still keep the source coding and channel coding separately, but optimize their parameters jointly [4].

In this paper, we propose an integrated framework to optimize the end-to-end performance of H.263 based video coding and transmission over wireless channels by considering robust source coding, channel coding and packetization, error concealment techniques together. For a given fixed channel signaling rate r and a family of channel code rate $r_c \in R_c$, the problem is to find the source and channel code pair (r_s, r_c) and corresponding video coding scheme such that the end-to-end distortion between original video signal at transmitter and video signal reconstructed at receiver is minimized. The end-to-end distortion is defined as

$$D_E(r_s, r_c) = \frac{1}{XYN} \sum_{x=1}^{X} \sum_{y=1}^{Y} \sum_{n=1}^{N} E\{[f_n^{(x,y)} - \tilde{f}_n^{(x,y)}(r_s, r_c)]^2\} \quad (1)$$

where X and Y are the dimension of the video frame, N is the total number of video frames. $f_n^{(x,y)}$ is the pixel value at spatial coordinate (x, y) in frame n of the original video signal. $\tilde{f}_n^{(x,y)}(r_s, r_c)$ is the reconstructed pixel value at (x, y) in frame n at the receiver using source and channel coding pair (r_s, r_c). The optimal pair (r_s^*, r_c^*) and corresponding video coding can be found as

$$(r_s^*, r_c^*) = \arg \min_{r_c \in R_c, r_s = r \cdot r_c} D_E(r_s, r_c) \quad (2)$$

For each pair (r_s, r_c), we assume that the H.263 based encoder knows the channel decoding failure rate when channel codeword is transmitted through noisy channels. The encoder determines the coding modes and quantizers for macroblocks in each frame using a rate-distortion optimized coding scheme to tradeoff between the source coding efficiency and robustness to error propagation. A recursive method which takes into account the inter-frame prediction and error propagation effect is used for the estimation of end-to-end distortion. Finally, the optimal source and channel code pair and corresponding robust video coding scheme can be found after the encoder performs the above procedure for all the finite number of source and channel code pairs. The optimal solution of minimizing (1) for each (r_s, r_c) is difficult to find because inter-frame prediction is used in video coding and error-propagation effects when errors occur. We use the greedy algorithm to find a near-optimal solution of (2) by looking for the optimal coding mode frame by frame during encoding.

In the following, we introduce the system parameters and a half-pel based distortion estimate algorithm which is used for rate-distortion optimized video coding scheme. Some simulation results are shown in Section 3. Finally, conclusions are reached.

2. ROBUST VIDEO CODING FOR A GIVEN SOURCE-CHANNEL RATE ALLOCATION

In this section, we describe the rate-distortion optimized video coding scheme for a source-channel rate pair (r_s, r_c). The system parameters such as channel coding and packetization are introduced first. Then a recursive distortion estimation scheme is introduced for half-pel based inter-frame video coding. Finally, rate-distortion based video coding scheme is described.

2.1. Channel Coding and Packetization at Encoder

Denote F the video frame rate and r the total channel bit rate, both of them are fixed. Base mode H.263 video coding standard and QCIF video sequence are used for simulation. There are total

N_s groups of macroblock (GOB) or slices in each Inter-frame, for QCIF format N_s is equal to 9. The GOB/slice structure is used to constrain the error propagation effects in the video bitstream where each GOB/slice is encoded independently. Reed-Solomon (RS) channel code over $GF(2^m)$ is used for FEC. There are N_s RS codewords used for each inter-frame in average. As a result, the number of codeword symbols in a $RS(n,k)$ codeword is

$$n = \frac{r}{F \cdot N_s \cdot m}. \tag{3}$$

For fixed RS codeword length n, now the problem is to find the optimal information symbol length k^* such that the corresponding end-to-end distortion in (1) is minimized.

To facilitate the error concealment at decoder when errors occur, the GOBs which are indexed by even numbers are concatenated and followed by concatenated GOBs indexed by odd numbers. Then after channel encoding, the RS codewords are transmitted over the wireless channels. By using this alternative GOB organization, the neighboring GOBs are not protected in the same RS codeword. Thus, when errors occur in one RS codeword, the neighboring GOBs will not be corrupted simultaneously, which helps the decoder to perform error concealment using the motion vectors of neighboring correctly received GOB.

For $RS(n,k)$, it can be calculated that the decoding failure probability $p_w(n,k)$ is $p_w(n,k) = \sum_{k=t+1}^{n} P(k,n)$, where $t = \lfloor (n-k)/2 \rfloor$ is the maximal number of symbol errors that can be corrected by $RS(n,k)$. $P(k,n)$ denotes the probability of k symbol errors within a block of n successive symbols which is known by the encoder. In order to estimate the end-to-end distortion, we need to derive the relation between video MB error probability P_{MB} and $p_w(n,k)$ of the channel code $RS(n,k)$, i.e.,

$$P_{MB}(n,k) = \alpha \cdot p_w(n,k) \tag{4}$$

It is difficult to find the exact relation between P_{MB} and $p_w(n,k)$ because the length of GOB in each frame is variable and each RS codeword dose not contain exactly one GOB. Simulations have been done and it turns out that $\alpha = 1.5$ is a good approximation in average. For a source and channel code pair (r_s, r_c), the channel code decoding failure probability $p_w(n,k)$ can be derived, then we have the corresponding MB error rate $P_{MB}(r_s, r_c)$ from (4). Based on the derived MB error rate $P_{MB}(r_s, r_c)$, a rate-distortion optimized video coding scheme is employed to tradeoff between the source coding efficiency and robustness to error propagation.

2.2. Recursive Estimate of End-to-end Distortion

The standard video coder such as H.263 and MPEG employs inter-frame prediction to remove temporal redundancies, and transform coding to exploit spatial redundancies. The video frame is segmented into MBs that are encoded either in Inter-mode or Intra-mode. Although Inter-mode generally achieves better compression, it incurs error propagation when errors occur. Recently, researches have been done to model the error propagation effects in order to optimally select mode for each MB to trade off the compression efficiency and error robustness [5][6][4].

Let $f_n^{\mathbf{s}}$ denote the original value of pixel at spatial location $\mathbf{s} = \{x, y\}$ in frame n and $\hat{f}_n^{\mathbf{s}}$ denote its reconstruction at encoder. The reconstructed value at the decoder is denoted by $\tilde{f}_n^{\mathbf{s}}$. From the encoder point of view, $\tilde{f}_n^{\mathbf{s}}$ is a random variable because of error propagation and concealment. The expected end-to-end distortion for pixel $\mathbf{s}$ in frame n is

$$d_n^{\mathbf{s}} = E\{(f_n^{\mathbf{s}} - \tilde{f}_n^{\mathbf{s}})^2\} = (f_n^{\mathbf{s}})^2 - 2 f_n^{\mathbf{s}} E\{\tilde{f}_n^{\mathbf{s}}\} + E\{(\tilde{f}_n^{\mathbf{s}})^2\} \tag{5}$$

In [5], recursive functions are derived to sequentially compute $E\{\tilde{f}_n^{\mathbf{s}}\}$ and $E\{(\tilde{f}_n^{\mathbf{s}})^2\}$ for integer-pel based video coding. For the half-pel case, the computation of spatial cross correlation between pixels in the same and different MBs are needed to obtain the first and second moment of bilinear interpolated half-pels, which is computationally very complex if not impossible. We circumvent this problem by using the error signals instead of pixel value.

Define $e_n^{\mathbf{s}} = f_n^{\mathbf{s}} - \hat{f}_n^{\mathbf{s}}$ the quantization error, $\hat{e}_n^{\mathbf{s},\mathbf{v}} = \hat{f}_n^{\mathbf{s}} - \hat{f}_{n-1}^{\mathbf{s}-\mathbf{v}}$ the prediction error using motion vector $\mathbf{v} = \{d_x, d_y\}$, and $\tilde{e}_n^{\mathbf{s}} = \hat{f}_n^{\mathbf{s}} - \tilde{f}_n^{\mathbf{s}}$ the transmission error. Assume $E\{\tilde{e}_n^{\mathbf{s}}\} = 0$, which is a reasonable assumption when P_{MB} is relatively low, we have

$$\begin{aligned} d_n^{\mathbf{s}} &= E\{(f_n^{\mathbf{s}} - \tilde{f}_n^{\mathbf{s}})^2\} \\ &= E\{(f_n^{\mathbf{s}} - \hat{f}_n^{\mathbf{s}} + \hat{f}_n^{\mathbf{s}} - \tilde{f}_n^{\mathbf{s}})^2\} \\ &= (e_n^{\mathbf{s}})^2 + E\{(\tilde{e}_n^{\mathbf{s}})^2\} \end{aligned} \tag{6}$$

We derive a recursive estimate of $E\{(\tilde{e}_n^{\mathbf{s}})^2\}$ for Intra-MB and Inter-MB. For Intra-mode MB, three cases are considered:

- with probability $1 - P_{MB}$, the Intra-MB is received correctly, then $f_n^{\mathbf{s}} = \tilde{f}_n^{\mathbf{s}}$. As a result, $\tilde{e}_n^{\mathbf{s}} = 0$.
- with probability $(1 - P_{MB})P_{MB}$, the Intra-MB is lost but the MB above is received correctly. Denoting $\mathbf{v}^c$ the motion vector of the MB above, we have $\tilde{f}_n^{\mathbf{s}} = \tilde{f}_{n-1}^{\mathbf{s}-\mathbf{v}^c}$ after error concealment. Then

$$\begin{aligned} \tilde{e}_n^{\mathbf{s}} &= \hat{f}_n^{\mathbf{s}} - \tilde{f}_n^{\mathbf{s}} = \hat{f}_n^{\mathbf{s}} - \tilde{f}_{n-1}^{\mathbf{s}-\mathbf{v}^c} \\ &= \hat{f}_n^{\mathbf{s}} - \hat{f}_{n-1}^{\mathbf{s}-\mathbf{v}^c} + \hat{f}_{n-1}^{\mathbf{s}-\mathbf{v}^c} - \tilde{f}_{n-1}^{\mathbf{s}-\mathbf{v}^c} \\ &= \hat{e}_n^{\mathbf{s},\mathbf{v}^c} + \tilde{e}_{n-1}^{\mathbf{s}-\mathbf{v}^c} \end{aligned} \tag{7}$$

- With probability P_{MB}^2, both current MB and previous MB are lost. The MB in previous frame at same location is repeated. We have

$$\tilde{e}_n^{\mathbf{s}} = \hat{f}_n^{\mathbf{s}} - \tilde{f}_n^{\mathbf{s}} = \hat{e}_n^{\mathbf{s},\mathbf{0}} + \tilde{e}_{n-1}^{\mathbf{s}} \tag{8}$$

finally we have

$$\begin{aligned} E_{Intra}\{(\tilde{e}_n^{\mathbf{s}})^2\} &= (1 - P_{MB})P_{MB}[(\hat{e}_n^{\mathbf{s},\mathbf{v}^c})^2 + E\{(\tilde{e}_{n-1}^{\mathbf{s}-\mathbf{v}^c})^2\}] \\ &+ P_{MB}^2[(\hat{e}_n^{\mathbf{s},\mathbf{0}})^2 + E\{(\tilde{e}_{n-1}^{\mathbf{s}})^2\}] \end{aligned} \tag{9}$$

Similarly, for Inter-mode coded MB we have

$$\begin{aligned} E_{Inter}\{(\tilde{e}_n^{\mathbf{s}})^2\} &= (1 - P_{MB})E\{(\tilde{e}_{n-1}^{\mathbf{s}-\mathbf{v}})^2\} \\ &\quad + (1 - P_{MB})P_{MB}[(\hat{e}_n^{\mathbf{s},\mathbf{v}^c})^2 + E\{(\tilde{e}_{n-1}^{\mathbf{s}-\mathbf{v}^c})^2\}] \\ &\quad + P_{MB}^2[(\hat{e}_n^{\mathbf{s},\mathbf{0}})^2 + E\{(\tilde{e}_{n-1}^{\mathbf{s}})^2\}] \end{aligned} \tag{10}$$

The encoder can use (9) and (10) to estimate the expected distotion $d_n^{(x,y)}$ recursively based on the accumulated coding and error-propagation effects from previous frames and current MB coding parameters.

2.3. Rate Distortion Optimized Video Coding

Rate-distortion optimized video coding is used to optimally select the quantizers and encoding modes of H.263 on the MB level for a specified MB error rate P_{MB} to tradeoff the source coding efficiency and robustness to error. We assume that a frame-level rate control algorithm has given us the maximal number of bits available R_{max} to code the current frame.

In H.263 video coding, there are $Row \times Col$ MBs in each video frame. For QCIF video sequences, $Row = 9$ and $Col = 11$. For each MB $b_{r,c}$, $r \in \{1, 2, \cdots, Row\}$ and $c \in \{1, 2, \cdots, Col\}$, Let $q_{r,c} \in Q$ be the quantizer parameter for $b_{r,c}$, where $Q = \{1, \cdots, 31\}$ is the set of admissible quantizer parameters. Let $m_{r,c} \in M$ be the encoding mode for $b_{r,c}$, where $M = \{Intra, Inter, Skip\}$ is the set of admissible encoding modes. Let $e_{r,c} = [q_{r,c}, m_{r,c}] \in E$ be the encoding vector for $b_{r,c}$, where $E = Q \times M$ is the set of all admissible decision vectors. Let $D(e_{r,c}, P_{MB})$ and $R(e_{r,c})$ be the distortion and the number of bits for $b_{r,c}$ when encoded using vector $e_{r,c}$.

Now the rate-distortion optimized video coding problem for Inter-frame n can be stated as

$$\min_{e_{r,c} \in E^{Row \times Col}} D_n(r_s, r_c) = \sum_{r=1}^{Row} \sum_{c=1}^{Col} D(e_{r,c}, P_{MB})$$

$$subject\ to \sum_{r=1}^{Row} \sum_{c=1}^{Col} R(e_{r,c}) \leq R_{max} \qquad (11)$$

where

$$D(e_{r,c}, P_{MB}) = \sum_{\mathbf{s} \in b_{r,c}} d^{\mathbf{s}} \qquad (12)$$

and $d^{\mathbf{s}}$ is computed using equation (6). This rate-distortion optimized video coding schemes have been studied for noiseless and noisy channels recently [7][8][9][6]. For video coding over error-prone channels, some simplifications have to be made: (a) GOB coding structure is used for H.263 video coding over noisy channels with each GOB is encoded independently. Therefore, if transmission errors occur in one GOB, the errors will not propagate into other GOBs in the same video frame. (b) The optimal decisions for each GOB have to be searched sequentially from top to bottom because the error concealment distortion for a corrupted MB is dependent on the motion vector of the MB above.

3. SIMULATION RESULTS

Simulations have been performed using the base mode H.263. In the simulations, the total channel signaling rate $r = 144$ kbps, which is a typical rate provided in the third generation wireless systems. Video frame rate is $F = 10$f/s. The video sequence used for simulation is *Foreman* in QCIF format. To simplify the simulation, a two-state Markov channel model [10] is used where the state transition is at symbol level. The two states of the model are denoted G (good) and B (bad). In state G symbols are received correctly whereas in state B symbols are erroneous. The model is fully described by the transition probabilities p_{GB} from state G to state B and p_{BG} from state B to state G. We use the probability of state B

$$P_B = \frac{p_{GB}}{p_{GB} + p_{BG}}, \qquad (13)$$

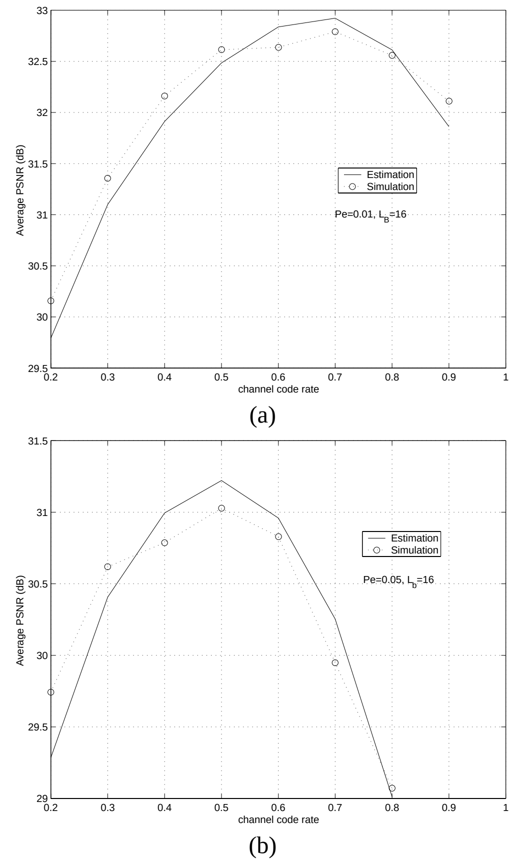


Fig. 1. Average PSNR obtained by estimation vs. simulation: (a) Symbol erro rate=0.01, (b) Symbol error rate=0.05.

and the average bursty length

$$L_B = \frac{1}{p_{BG}}. \qquad (14)$$

which is the average number of consecutive symbol errors to model the two-state Markov model.

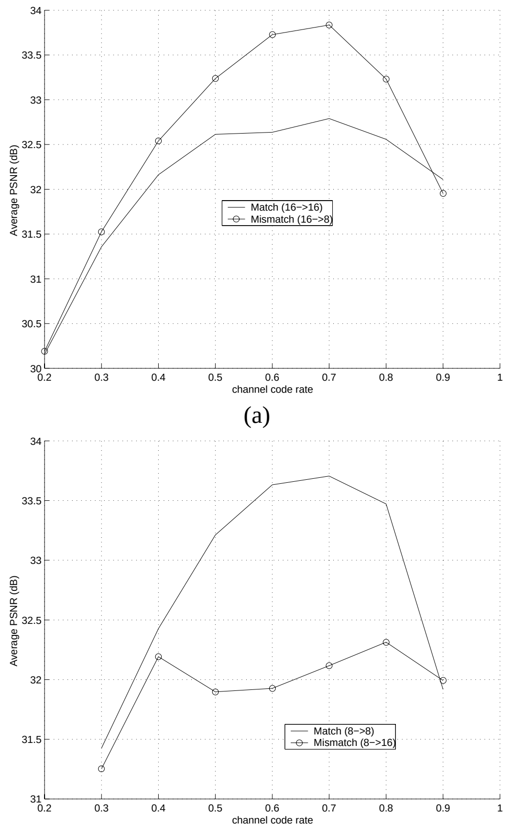
RS code over GF(2^8) is used for FEC. The average peak signal-to-noise ratio (PSNR) $PSNR_E$ is used to measure the performance

$$PSNR_E(r_s, r_c) = \frac{1}{N} \sum_{n=1}^{N} PSNR_E^n(r_s, r_c) \qquad (15)$$

where $PSNR_E^n(r_s, r_c)$ is the estimated average PSNR between original frame n and corresponding reconstruction at decoder using pair (r_s, r_c).

Fig. 1(a) shows the average estimated $PSNR_E$ of the video coding after optimal rate allocation and robust video coding for different channel code rate when the symbol error rate $P_B = 0.01$ and bursty length $L_B = 16$ symbols, and the corresponding simulated average $PSNR_S$ of 50 times video transmission. The $PSNR_S$ is defined as

$$PSNR_S(r_s, r_c) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{N} \sum_{n=1}^{N} PSNR_S^{(n,l)}(r_s, r_c) \qquad (16)$$



(a)

(b)

Fig. 2. Average PSNR obtained in channel mismatch cases:(a) bursty is shorter then that used in estimation, (b) bursty is longer then that used in estimation.

where $PSNR_S^{(n,l)}(r_s, r_c)$ is the PSNR between original frame n and corresponding reconstruction at decoder in lth simulation using pair (r_s, r_c). Fig. 1(b) also shows the same comparison when the symbol error rate $P_B = 0.05$. It can be noted that the estimated PSNR, which is obtained at encoder during video encoding, matches the simulated PSNR very well. The optimal source coder and channel coder pair can also be found through Fig. 1(a) and (b) for different channel characteristics. The corresponding channel decoding failure rate of the optimal channel coding rates in Fig. 1(a) and (b) are 0.018 and 0.034, respectively.

We also compared the performance when the channel model used at video encoder does not match the real channel characteristics used in simulations. Fig.2 show two cases of channel mismatch. In Fig.2(a), video stream which is encoded based on $P_B = 0.01, L_B = 16$ two-state Markov channel is simulated using two-state Markov channel with $P_B = 0.01, L_B = 8$. The simulated average PSNR is better than the average PSNR estimated at encoder during encoding. On the other hand, when the video stream which is encoded based on $PB = 0.01, L_B = 8$ two-state Markov channel is simulated using two-state Markov channel with $P_B = 0.01, L_B = 16$, the simulated average PSNR is much worse than the average PSNR estimated at encoder as shown in Fig.2(b). Further more, the optimal source and channel coder pair obtained at encoder is not optimal when the channel condition is worse than the channel information used at the encoder. This simulation result suggests that the optimal rate allocation and video coding should be focused on the worse channel conditions for broadcasting services.

4. CONCLUSIONS

We have proposed an integrated framework to optimally allocate the total channel rate to the H.263 video source coder and channel coder and obtain the corresponding robust video coding scheme for transmission over wireless channels when there is no feedback channel available. Assuming the encoder has the stochastic channel information of the wireless fading channel, the proposed scheme takes into account the robust video coding, packetization, error propagation and concealment effects together. Simulation results demonstrated the optimality of the rate allocation scheme and accuracy of the estimated distortion obtained during the process of video encoding at encoder.

5. REFERENCES

[1] J. Modestino and D. G. Daut, “Combined source-channel coding of images,” *IEEE Trans. Communications*, vol. COM-27, pp. pp.1644–1659, Nov. 1979.

[2] N. Tanabe and N. Farvardin, “Subband image coding using entropy coded quantization over noisy channels,” *IEEE J. Select. Areas Commun.*, vol. 10, pp. 926–943, June 1992.

[3] H. Jafarkhani, P. Ligdas, and N. Farvardin, “Adaptive rate allocation in a joint source/channel coding framework for wireless channels,” *Proceedings of Vehicular Technology Conference*, pp. 492–496, 1996.

[4] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, “Analysis of video transmission over lossy channels,” *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1012–1032, June 2000.

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

[6] G. Cote, S. Shrani, and F. Kossentini, “Robust h.263 video communication over mobile channels,” *IEEE Proceedings of International Conference on Image Processing*, June 1999.

[7] K. Ramchandran, A. Ortega, and M. Vetterli, “Bit allocation for dependent quantization with applications to multiresolution and mpeg video coder,” *IEEE Trans. Image Processing*, vol. 3, pp. 533–545, Sep. 1994.

[8] G. M. Schuster and A. K. Katsaggelos, “Fast and efficient mode and quantizer selection in the rate distortion sense for h.263,” *Proceedings of SPIE*, vol. 2727, pp. 784–795, 1996.

[9] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. Mitra, “Rate-distortion optimized mode selection for very low bit rate video coding and the emerging h.263 standard,” *IEEE Trans. on Circuits and Syst. for Video Technology*, vol. 6, pp. 182–190, April 1996.

[10] E. O. Elliott, “Estimates of error rate for codes on burst-noise channels,” *Bell Syst. Tech. J.*, vol. 42, pp. 1977–1997, Sep. 1963.