

OPTIMAL MODE SELECTION AND CHANNEL CODING FOR VIDEO TRANSMISSION OVER WIRELESS CHANNELS USING H.264/AVC

*Ehsan Maani*¹, *Fan Zhai*², and *Aggelos K. Katsaggelos*¹

¹Northwestern University, Evanston, IL 60208, USA

²Texas Instruments, Dallas, TX 75243, USA

ABSTRACT

This paper addresses the problem of joint encoder optimization and channel coding for realtime video transmission over wireless channels. An efficient solution is proposed to optimally select macroblock modes and quantizers as well as channel coding rates. The proposed optimization algorithm fully considers error resilience, forward error correction and error concealment. Experimental results demonstrate the effectiveness of the proposed approach.

Index Terms— Error concealment, error control, error resilience, multimedia streaming, quality of service (QoS), resource allocation, unequal error protection (UEP).

1. INTRODUCTION

With the increasing bandwidth in the next generations of mobile networks and the rapidly growing demand for video applications, such as video conferencing, distant learning, and on-demand video streaming, wireless video transmission has received much attention during the last few years. Due to the limited and time-varying channel bandwidth, video transmission over wireless channels is still a challenging task. In order to efficiently utilize the limited bandwidth, video signals have to be highly compressed by efficient coding algorithms; however, highly compressed video data becomes extremely vulnerable to bit corruption [1]. Therefore, it is a requirement for the video encoder to provide adequate error resilience features to protect video data from channel errors. In general, error control techniques include error resilient source coding, forward error correction (FEC), Automatic Repeat request (ARQ), power adjustment and network QoS support [2]. The problem of Joint Source Channel Coding (JSCC) jointly considers allocation of the available error control components across layers in a rate-distortion optimization framework. JSCC using FEC has been extensively studied in the literature. In [3], [4], [5], and [6], authors present their work in context of wireless channels for scalable video transmission. In these papers, error resilience is achieved through optimized transport prioritization for layered video. In [7], JSCC with error resilient source coding and FEC for Internet scalable video is studied. A rate-distortion optimized mode selection and synchronization marker insertion algorithm is introduced in [8] as an effective method to increase source error resilience. However, in these works, not all possible error control components are considered simultaneously at the application layer. In a more recent work [9], error resilient source coding, channel coding and error concealment are addressed in an integrated optimization framework where the interaction between source coding and channel coding is fully considered. Nevertheless, in this framework, all macroblocks in a packet are considered to have the same prediction mode and quantizer, which may degrade the source compression efficiency.

In this paper we expand the work in [9] removing the last constraint while H.264/AVC is employed instead of H.263+. We propose an algorithm for designing a real-time video transmission system that optimally selects prediction mode and quantizer per macroblock, as well as, per packet channel protection rate, while error concealment is being taken into account. The rest of this paper is organized as follows. Section 2 provides some basic preliminaries that will be used throughout the paper. The mathematical formulation of the problem is provided in Section 3. Section 4 discusses the solution algorithm. Section 5 presents some experimental results, and section 6 concludes the paper.

2. BACKGROUND

2.1. System Model

In a typical packet-based video communication system, video data is encoded using a block-based motion-compensated technique where each frame is divided into slices that are comprised of consecutive macro-blocks (MBs). Each slice is independently decodable, i.e., the decoding of one slice is not affected by the loss of other slices in the same frame. Losses in other frames, on the other hand, may cause temporal error propagation due to inter-frame prediction. Here, in our formulations, each slice is considered to be one packet i.e., a row of macroblocks.

After each packet is encoded, parity check bits are generated and added to the packet and then the it is transmitted over the wireless link. At the receiver, the packet is considered lost when an unrecoverable number of bit errors occurs. In our simulations, we use a simple Gilbert-Elliott two-state channel model. In addition, for performing link-layer FEC, the use of rate-compatible punctured convolutional (RCPC) codes is considered [10].

2.2. Error Concealment

To avoid significant visible errors in the reconstructed frames at the decoder, the implementation of a robust error concealment technique is inevitable. Here, to perform error concealment we use an algorithm similar to the one implemented in H.26L test model [11]. In this algorithm, all correctly received slices of a picture are decoded first, and then the lost slices are concealed. A record of the status of each macroblock is kept, where each MB can be one of the following, “Correctly received”, “Lost” and “Concealed”. Since, “Concealed” MBs are also treated as reliable in the concealment process, the order in which “Lost” MBs are concealed is also important. The concealment starts by processing MB columns at the frame boundaries and then moves inwards MB column by MB column. The algorithm for inter-coded frames tries to estimate the motion in the missing MB by prediction schemes from available motion information

of spatial neighbors. In this work, since causality plays an important role in our optimization algorithm, the information of the MBs below the “Lost” MB is not used for concealment.

2.3. End-to-End Expected Distortion

Due to channel losses, we use the expected end-to-end distortion to evaluate video quality. Three factors can be identified as effecting the end-to-end distortion: the source behavior (quantization and packetization), the channel characteristics, and the receiver behavior (error concealment) [12, 13]. The expected distortion can be calculated at the encoder as

$$E[D_k] = (1 - p_k)E[D_{R,k}] + p_k(1 - p_{k-1})E[D_{LR,k}] + p_k p_{k-1}E[D_{LL,k}], \quad (1)$$

where p_k is the loss probability for the k^{th} packet, $E[D_{R,k}]$ is the expected distortion if the k^{th} packet is received, and $E[D_{LR,k}]$ and $E[D_{LL,k}]$ are respectively the expected distortion after concealment when packet $(k - 1)$ is either received or lost. The evaluation of the expected distortion is based on a per pixel recursive algorithm called ROPE [14]. The accuracy of ROPE in end-to-end distortion estimation is attributed to its ability to calculate the first and second moments of the reconstructed pixels by the decoder; however, sub-pixel prediction employed in H.264 involves interpolation of neighboring pixels, which gives rise to cross-correlation terms in the second moment calculation. To deal with the cross correlation terms in our experiments, the cross correlation approximation method introduced in [15] is used.

3. PROBLEM FORMULATION

The following notation will be used throughout this paper. Let $m_{r,c}$ be the macroblock at row r and column c , where $r \in [1, \dots, M]$ and $c \in [1, \dots, N]$. Let $\mu_{r,c} \in \mathcal{Q}$ denote the source parameter for macroblock $m_{r,c}$ where \mathcal{Q} is the set of all admissible MB modes with different quantization parameters, i.e., $\mathcal{Q} = \{intra, inter, skip\}$, where *intra* and *inter* modes include the quantization parameter set $\{qp, qp \pm 2, qp \pm 4\}$ for a given parameter qp . The *skip* mode is a special case of the inter mode where no information is transmitted, and the block is simply repeated from the spatially corresponding block in the previous frame. Furthermore, let \mathcal{R} represent the set of all channel coding parameters, i.e., FEC with different rates. Finally, let ν_r represent the channel coding rate of the r^{th} row which is contained in one packet.

Our goal is to find modes, quantizers and channel coding rates that minimize the expected decoder distortion. Furthermore, each packet has to meet a delay constraint in order to reach the receiver in time for playback. If F denotes the playback rate at the receiver, the transmission time T for each frame has to be less than $T_0 = 1/F$. Consequently, the joint source channel problem can be formulated as the following constrained optimization problem

$$\begin{aligned} \min_{\nu \in \mathcal{R}^M, \mu \in \mathcal{Q}^M \otimes \mathcal{Q}^N} & \sum_{r=1}^M E[D_r(\mu_r, \mu_{r-1}, \nu_r, \nu_{r-1})] \\ s.t. & T = \sum_{r=1}^M R_r(\mu_r, \nu_r) / R_T \leq T_0, \end{aligned} \quad (2)$$

where, the symbol \otimes is the direct product of two sets and $R_r(\mu_r, \nu_r)$ is total number of bits allocated to the r^{th} row after channel coding,

which is a function of source coding parameters μ_r and channel coding rate ν_r used for that packet. Note that due to error concealment, the expected distortion of the r^{th} packet D_r , also depends on the source parameters and channel coding rate of the preceding packet $(r - 1)$.

4. SOLUTION ALGORITHM

In this section, we present the solution algorithm based on the concept of the Lagrangian relaxation. The Lagrangian, with a multiplier λ , can be expressed as

$$\mathcal{L}(\mu_r, \mu_{r-1}, \nu_r, \nu_{r-1}; \lambda) = E[D] + \lambda(T - T_0), \quad (3)$$

where $E[D]$ and R are the total per frame expected distortion and rate, respectively. Furthermore, the dual function to this optimization problem can be defined as

$$q(\lambda) = \min_{\nu \in \mathcal{R}^M, \mu \in \mathcal{Q}^M \otimes \mathcal{Q}^N} \mathcal{L}(\mu_r, \mu_{r-1}, \nu_r, \nu_{r-1}; \lambda). \quad (4)$$

The solution to the formulation (2) within a convex hull approximation is obtained by

$$\max_{\lambda \geq 0} q(\lambda). \quad (5)$$

Assuming that we can evaluate the dual function (4) for a given choice of λ , a solution to the JSCC problem can be found by choosing the correct Lagrange multiplier λ^* that satisfies the rate constraint in (2). This can be accomplished by using a variety of methods such as bisection, cutting-plane or sub-gradient methods [16]. To evaluate the dual function (4), Dynamic Programming (DP) can be employed. This process is discussed in detail in the next section.

4.1. Minimization of Lagrangian

Evaluation of the dual function (4) for a given Lagrange multiplier is still a complicated task due to inter and intra packet dependencies. The minimization of the Lagrangian for a given multiplier is therefore carried out in two steps. First, in each row, optimum macroblock modes are selected for any channel parameter; then, using the Viterbi algorithm, the channel coding rate with the lowest cost is assigned to each packet.

4.1.1. Optimal Mode Selection

For the moment let us assume that the channel parameters for the current packet r and previous packet $r - 1$, $(\nu_r, \nu_{r-1}) \in \mathcal{R} \times \mathcal{R}$, are given. Due to the specific selected GOB structure, the predictors for the current motion vectors are just the motion vectors from the macroblock immediately to the left. Therefore, the rate to encode the current macroblock $m_{r,c}$, $R_{r,c}(\mu_{r,c-1}, \mu_{r,c})$, depends on the mode and quantizer choices for $m_{r,c-1}$ and $m_{r,c}$. Moreover, in principle, reconstructed pixel values of a macroblock encoded by H.264 depend on the left neighboring MB due to different prediction schemes. Nevertheless, in this work since the sub-macroblock rate distortion optimization is performed independently from any of the neighboring macroblocks, it is assumed that the encoded pixel values do not depend on the selected modes for the neighboring macroblocks. As a result, the macroblock expected distortion $D_{r,c}(\mu_{r,c}, \nu_r, \nu_{r-1})$ only depends on the current MB mode and channel parameters. Hence, the row expected distortion can be expressed as

$$D_r(\mu_r, \nu_r, \nu_{r-1}) = \sum_{c=0}^N D_{r,c}(\mu_{r,c}, \nu_r, \nu_{r-1}), \quad (6)$$

and the row rate can be expressed as a sum over the macroblock rates,

$$R_r(\boldsymbol{\mu}_r, \nu_r) = \sum_{c=0}^N R_{r,c}(\mu_{r,c-1}, \mu_{r,c}) \cdot (1/\nu_r). \quad (7)$$

In this scenario, to find the optimal MB modes a similar method to [17] is employed. Depending on the modes $\mu_{r,c-1}$ and $\mu_{r,c}$ one of the following predictions may be used to calculate the macroblock rate $R_{r,c}(\mu_{r,c-1}, \mu_{r,c})$

(1) *Intra Prediction*: In H.264/AVC intra prediction is conducted in the spatial domain, by referring to neighboring samples of previously coded blocks which are to the left and/or above the block to be predicted. In this work, to prevent error propagation due to motion compensation, intra prediction is only restricted to predict from intra-coded neighboring macroblocks. Therefore, when both $m_{r,c-1}$ and $m_{r,c}$ are to be intra-coded, the intra prediction process needs to be invoked to calculate the intra-predicted rate $R_{r,c}^{IP}(\mu_{r,c-1}, \mu_{r,c})$. However, to avoid heavy calculations of the rate $R_{r,c}^{IP}$ for every quantization pair $(\mu_{r,c-1}, \mu_{r,c})$, the bit saving ratio due to intra prediction η is calculated for the average quantizer $\bar{\mu}_{r,c} \in \mathcal{Q}$ as

$$\eta = R_{r,c}^{IP}(\bar{\mu}_{r,c-1}, \bar{\mu}_{r,c}) / R_{r,c}(\bar{\mu}_{r,c-1}, \bar{\mu}_{r,c}), \quad (8)$$

where, $R_{r,c}$ is the rate without intra prediction. Then the ratio η is used to approximate the intra predicted rate for other quantization parameters of the current macroblock. It is assumed that a change in the quantizer parameter of the previous macroblock has a negligible effect on the rate $R_{r,c}^{IP}$. Our simulations show that the average error in the MB rate due to this approximation is less than 4%.

(2) *Inter Prediction*: In the case of an inter mode, the macroblock rate $R_{r,c}(\mu_{r,c-1}, \mu_{r,c})$ depends on the following three different terms,

$$R_{r,c}(\mu_{r,c-1}, \mu_{r,c}) = R_{r,c}^{DFD}(\mu_{r,c}) + R_{r,c}^{DVF}(\mu_{r,c-1}, \mu_{r,c}) + R_{r,c}^Q(\mu_{r,c-1}, \mu_{r,c}), \quad (9)$$

(1) The rate to encode the prediction error or displaced frame difference $R_{r,c}^{DFD}$ which only depends on the quantizer and mode selection for the current macroblock $m_{r,c}$. (2) The rate to encode the displacement vector field $R_{r,c}^{DVF}$, which depends on modes and quantizers of both macroblocks $m_{r,c-1}$ and $m_{r,c}$. Note that with the change of the quantizer, motion vectors and macroblock type might change due to the sub-macroblock, quantizer dependent rate-distortion optimization of the H.264/AVC. (3) The rate to transmit the quantizer difference of $\mu_{r,c-1}$ and $\mu_{r,c}$, which obviously depends on both quantizers.

Using the above macroblock rates and distortion, the Lagrangian cost function for a macroblock can be expressed as

$$J_{r,c}(\mu_{r,c-1}, \mu_{r,c}, \nu_{r-1}, \nu_r; \lambda) = D_{r,c}(\mu_{r,c}, \nu_{r-1}, \nu_r) + \lambda R_{r,c}(\mu_{r,c-1}, \mu_{r,c}, \nu_r) \quad (10)$$

Therefore, the optimal mode selection can be written as

$$J_r(\nu_{r-1}, \nu_r) = \min_{\boldsymbol{\mu}_r \in \mathcal{Q}^N} \sum_{c=0}^N J_{r,c}(\mu_{r,c-1}, \mu_{r,c}, \nu_{r-1}, \nu_r; \lambda) \quad (11)$$

The above problem can be efficiently solved using Dynamic Programming (DP), which is also known as the Viterbi algorithm.

4.1.2. Optimal Channel Coding

In the previous section we discussed how we can efficiently employ the Viterbi algorithm to optimally select MB modes for any pair of channel parameters $(\nu_r, \nu_{r-1}) \in \mathcal{R} \times \mathcal{R}$. As a result, the Lagrangian minimization (4) simplifies to

$$q(\lambda) = \min_{\nu \in \mathcal{R}, \mathcal{M}} \sum_{r=0}^M J_r(\nu_{r-1}, \nu_r), \quad (12)$$

where, $J_r(\nu_{r-1}, \nu_r)$ is given by equation (11). Dynamic Programming can be employed to efficiently solve this minimization problem. DP can be viewed as the shortest path problem in a trellis, where each stage corresponds to the channel rate selection for a given packet. Recall that since MB modes are already selected for each channel rate pair, rate and expected distortion of the row is known, thus, the total cost of the packet can be easily evaluated.

5. EXPERIMENTAL RESULTS

Our simulations are based on the modified H.264/AVC JM 10.2 software, where the IPPP...P structure, one reference frame, 16 pixel search range for ME/MC, and CAVLC are enabled. We use an RCPC code with memory $M = 4$, period $P = 8$, and rates $R = P/(P + l)$, $l = 2, 4, \dots, 12$. At the receiver, soft Viterbi decoding is used in conjunction with BPSK demodulation. As stated before, the source parameter set for the proposed work is chosen to be $\mathcal{Q} = \{\textit{intra}, \textit{inter}, \textit{skip}\}$ with quantizers $qp, qp \pm 2, qp \pm 4$ for *intra* and *inter* modes. Here, qp is a fixed parameter in our optimization that depends on the sequence resolution and transmission rate (in this case $qp = 30$). The channel is modeled by a two-state Markov chain with state space $\mathcal{H} = \{0.05, 0.005\}$. The transition probabilities $T_{1 \rightarrow 2}$ and $T_{2 \rightarrow 1}$ are 0.12 and 0.06 respectively.

Here, we compare three different transmission systems. The first system uses the proposed framework with the parameters explained above. The second system employs the Integrated Joint Source Channel Coding algorithm in [9] to perform optimization while other parameters are the same as the first system. In this system one prediction mode and quantizer is assumed per row of MBs. Finally, the third system uses the built-in H.264 rate control for source coding with a fixed channel coding rate that is the average of the channel rates used in system 1. Error resilience is incorporated in this system by means of random intra MB refreshing.

Figure 1 illustrates the performance of all three systems when a Markov channel model is used. It is assumed that the encoder does not have any prior information about the status of the lossy channel. However, channel feedback (packet ACK/NACK) is available every 66 ms. To be able to explain the large gap in performance of system 1 and the other systems, let us consider a situation where some packets are lost in the channel due to deep fading. System 1 will optimally allocate intra macroblocks in the area that is mostly distorted to prevent error propagation in future frames. On the contrary, system 2 only optimizes over packet (a row of MBs) modes and intra-coded packets can be quite expensive in terms of bits. System 3 randomly allocates intra MBs in the frame. As a result, the proposed algorithm (system 1) can be very effective in terms of error recovery and therefore has a higher average PSNR.

6. CONCLUSION

In this paper, we studied the joint source and channel optimization problem for real-time video transmission over lossy channels. In

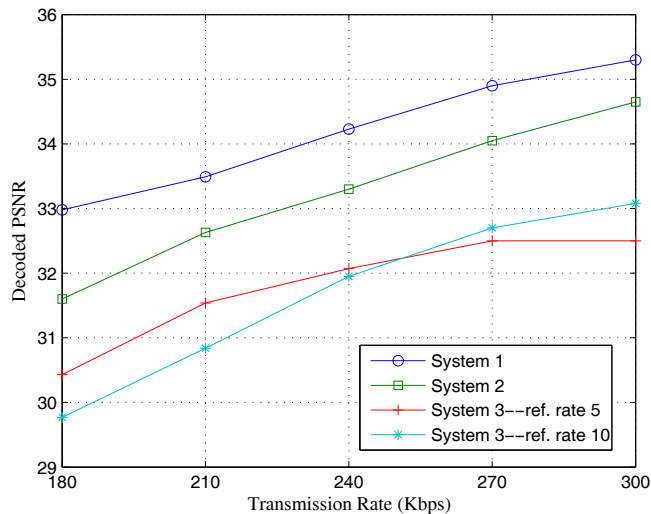


Fig. 1. Average PSNR comparison of the three systems

the proposed algorithm, optimal MB modes and associated parameters are jointly selected with the channel coding rate. Experimental results show significant improvement in quality of the transmitted video.

7. REFERENCES

- [1] K. Stuhlmüller, N. Farber, Mlink, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1012–1032, June 2000.
- [2] A. K. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry, and T. N. Pappas, "Advances in efficient resource allocation for packet-based real-time video transmission," *Proc. IEEE*, vol. 93, no. 1, Jan. 2005.
- [3] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "Network-adaptive scalable video streaming over 3G wireless network," in *Proc. IEEE Int. Conf. Image Processing*, Thessaloniki, Greece, Oct. 2001, pp. 579–582.
- [4] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 511–523, June 2002.
- [5] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint source-channel coding for motion compensated DCT-based SNR scalable video," *IEEE Trans. Image Processing*, vol. 11, pp. 1043–1052, Sept. 2002.
- [6] S. Appadwedula, D. L. Jones, K. Ramchandran, and L. Qian, "Joint source-channel matching for wireless image transmission," in *IEEE Int. Conf. on Image Processing*, Chicago, IL, Oct. 1998.
- [7] M. Gallant and F. Kossentini, "Rate-distortion optimized layered coding with unequal error protection for robust internet video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 357–372, Mar. 2001.
- [8] G. Gôte, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 952–965, June 2002.
- [9] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Rate-distortion optimized hybrid error control for packetized video communications," *IEEE Trans. Image Processing*, vol. 15, pp. 40–53, Jan. 2006.
- [10] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, pp. 389–400, Apr. 1988.
- [11] Y. Wang, M. M. Hannuksela, V. Varsa, A. Hourunranta, and M. Gabbouj, "The error concealment feature in the H.26L test model," in *Proc. ICIP*, Rochester, NY, Sept. 2002, pp. 729–732.
- [12] D. Wu, Y. T. Hou, B. Li, W. Zhu, Y. Q. Zhang, and H. J. Chao, "An end-to-end approach for optimal mode selection in internet video communication: theory and application," *IEEE Trans. Commun.*, vol. 18, pp. 977–995, June 2002.
- [13] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source-channel coding and power allocation for energy efficient wireless video communications," in *Proceedings of the 41st Allerton Conference on Communications, Control, and Computing*, Monticello, IL, Oct. 2003.
- [14] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Trans. Commun.*, vol. 18, pp. 966–976, June 2000.
- [15] H. Yang and K. Rose, "Advances in recursive per-pixel estimation of end-to-end distortion for application in H.264," in *Proc. ICIP*, Genova, Sept. 2005.
- [16] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, second edition, 1999.
- [17] G. M. Schuster and A. K. Katsaggelos, "Fast and efficient mode and quantizer selection in the rate distortion sense for H.263," Orlando, FL, Mar. 1996.