AN INTEGRATED JOINT SOURCE-CHANNEL CODING FRAMEWORK FOR VIDEO TRANSMISSION OVER PACKET LOSSY NETWORKS

Fan Zhai, Yiftach Eisenberg, Thrasyvoulos N. Pappas,
Randall Berry, and Aggelos K. Katsaggelos

Department of Electrical and Computer Engineering
Northwestern University, Evanston, IL 60208, USA
E-mail: {fzhai, yeisenbe, pappas, rberry, aggk}@ece.northwestern.edu

ABSTRACT

The problem of application-layer error control for real-time video transmission over packet lossy networks is commonly addressed by joint source-channel coding (JSCC). The traditional JSCC approaches solve this problem in a sequential manner, where source coding and channel coding are not fully integrated. In this paper, we present an integrated joint source-channel coding (IJSCC) framework, where error resilient source coding, channel coding, and error concealment are jointly considered in an integrated manner. We show through both analysis and simulations the advantages of the proposed IJSCC approach, in comparison to a sequential JSCC approach.

1. INTRODUCTION

Real-time video applications, such as on-demand video streaming, videophone and videoconferencing, have gained increased popularity. However, it is well known that the best effort design of the current Internet makes it difficult to provide the quality of service (QoS) needed by these applications. A direct approach for dealing with the lack of QoS is to use error control. In this paper, we consider a combination of common error control approaches. Specifically, we consider error resilient source coding and error correction at the sender side, and error concealment at the receiver. We present an integrated joint source channel coding (IJSCC) framework for jointly optimizing these application-layer error control components to achieve the best video quality.

Each of the above error control approaches is designed to deal with a lossy packet channel. Error resilient source coding accomplishes this by adding redundancy at the source coding level to prevent error propagation and limit the distortion caused by packet losses. Another approach is to use error correction techniques in the application/transport layer. Two basic techniques are commonly used: Forward Error Correction (FEC) and Automatic Repeat reQuest (ARQ). Of the two error correction techniques, FEC-based techniques are usually preferred for video applications and are currently under consideration by the IETF as a proposed standard in supporting error resilience [1]. This is mainly because ARQ cannot accommodate the delay requirements of real-time video applications. Finally, error concealment refers to post-processing techniques employed by the decoder to recover from packet loss by utilizing the spatial and temporal correlation of the video sequence [2].

Error control for video transmission is often studied in a JSCC framework, e.g., [3–8]. JSCC has three tasks: finding an optimal bit allocation between source coding and channel coding for given channel loss characteristics; designing the source coding to achieve the target source rate; and designing the channel coding to achieve the required robustness [2].

Most of the JSCC work to date has focused on bit allocation between source and channel coding, such as in [3, 5–8]. Source coding is performed based on the given bit budget, after the bit allocation between source and channel is completed. The optimization of source coding can be achieved in the form of mode selection by taking into account the residual errors after channel coding, such as in [4, 9]. The above studies, however, do not fully consider the interaction between source coding and channel coding. More specifically, they do not take into account the effect of error resilient source coding upon the bit allocation between source and channel. In this work, we introduce the IJSCC framework, where error resilient source coding, channel coding, and error concealment are jointly considered in a tractable optimization setting. This framework has been employed in [10, 11]

2. SEQUENTIAL JOINT SOURCE-CHANNEL CODING

Let $\mathcal{Q}$ and $\mathcal{R}$ be the set of source coding parameters and FEC parameters, respectively. The source coding parameters typically include the prediction mode and quantization step size. The FEC parameters represent the amount of overhead devoted to error resilience. Let $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ denote respectively the vector of source coding parameters and channel coding parameters for one frame. Let the superscript (n) denote the frame index, and the subscripts s and c stand for source and channel coding, respectively. Then, the sequential two-step JSCC can be formally represented as

$$\begin{aligned} &\min_{\{\boldsymbol{\nu}\in\mathcal{R}\}} E[D^{(n)}(\boldsymbol{\nu})] \\ \text{s.t. } T^{(n)}(\boldsymbol{\nu}) =& B_s^{(n)}(\boldsymbol{\mu}(\boldsymbol{\nu}))/R_T + B_c^{(n)}(\boldsymbol{\nu})/R_T \le T_0^{(n)}, \end{aligned} \tag{1}$$

and

$$\begin{aligned} &\min_{\{\boldsymbol{\mu}\in\mathcal{Q}\}} E[D^{(n)}(\boldsymbol{\mu})] \\ &\text{s.t. } T_s^{(n)}(\boldsymbol{\mu}) = B_s^{(n)}(\boldsymbol{\mu})/R_T \le T_{s,0}^{(n)}, \end{aligned} \tag{2}$$

where $E[D]$ is the expected distortion, R_T the transmission rate, B_s and B_c the source bits and channel bits, respectively, T the associated transmission delay, and T_0 and $T_{s,0}$ the transmission delay constraints for the whole frame (including both source and channel bits) and the source bits, respectively. In (1), the constraint

is on the total transmission delay for the n-th frame, $T^{(n)}$; in (2), the constraint is on the source transmission delay[1], $T_s^{(n)}$. Several channel coding techniques have been considered for solving (1). For work utilizing pre-encoded video, such as [3], source coding is fixed. Thus, the objective is to minimize the channel induced distortion, and the second step (2) is not necessary. For work on coding the source on the fly, one way to characterize the distortion in (1) is to use a source R-D model, as in [5,6]. For example, a universal RD model is used in [6]. In [5], the distortion is expressed as the sum of source and channel distortion, both of which are model-based. By assuming uncorrelated source and channel distortion, the first-step of the minimization in [5] aims at minimizing the channel distortion, while the second-step minimizes the source distortion. There has also been considerable work in the area of JSCC for wavelet-based scalable video coders, such as [8]. The inherent prioritization of information in a wavelet-based video bitstream makes the implementation of JSCC rather straightforward. For block-based motion compensated video coding, JSCC is more challenging because the relative importance of packets is not explicitly available.

The above studies, however, do not fully consider the interaction between source coding and channel coding. The goal of JSCC is to compress the source in an error resilient manner and add redundancy through the channel code to achieve the best trade-off between error robustness and compression efficiency. The optimal way to achieve this requires the joint consideration of error resilient source coding and channel coding. It is clear that such an integrated approach should be superior to the sequential approach in (1) and (2).

3. INTEGRATED JOINT SOURCE-CHANNEL CODING

In our IJSCC framework, instead of separating the overall expected distortion into source distortion and channel distortion, as in (1) and (2), we consider the interaction between these components. This approach is based on the fact that for optimal results, besides FEC, which adapts to channel characteristics before error recovery, source coding should also be adapted to the modified channel characteristics after error recovery.

A related framework was presented in [4] for jointly considering error resilient source coding and channel coding. In that work, the distortion measurement was model-based. Here, we recursively calculate packet distortion, which takes into account both source distortion and channel distortion, as well as error propagation due to channel errors[2]. Our objective is to minimize the total expected distortion for the n-th frame, given a transmission delay constraint, i.e.,

$$\begin{aligned} &\min_{\{\boldsymbol{\mu}\in\mathcal{Q},\boldsymbol{\nu}\in\mathcal{R}\}} E[D^{(n)}(\boldsymbol{\mu},\boldsymbol{\nu})] \\ &\text{s.t.} \quad T^{(n)}(\boldsymbol{\mu},\boldsymbol{\nu}) = B^{(n)}(\boldsymbol{\mu},\boldsymbol{\nu})/R_T \le T_0^{(n)}, \end{aligned} \tag{3}$$

where $B^{(n)}$ represents the total bits used for both source coding and channel coding, and $T_0^{(n)}$ is the transmission delay constraint for this frame.

[1]Note both of these constraints can also be interpreted as specifying bit budgets of $T_0^{(n)}R_T$ and $T_{s,0}^{(n)}R_T$.

[2]The effect of error propagation can be fully captured based on the acknowledgement information after 1 round-trip-time (RTT) delay.

4. SIMULATION ISSUES

4.1. Channel Model

The proposed framework is general and not limited to any specific packet loss model. All that is needed is a stochastic model of the packet losses. For simplicity, in this paper, packet loss in the network is modeled by a Bernoulli process, i.e., each packet is independently lost with probability ϵ. We assume that the receiver responds to a lost or corrupt packet with a negative acknowledgement, and responds to a correctly received packet with a positive acknowledgement. All acknowledgements are assumed to arrive correctly after one RTT, i.e., the feedback delay is a constant and the feedback channel is assumed to be error free.

4.2. Packetization and Error Concealment

We consider a system where each row of blocks is coded as one source packet and every packet is independently decoded. We employ Reed-Solomon coding, a widely used erasure code, to provide inter-packet FEC as in [4, 10]. For error concealment, we consider a temporal replacement error concealment strategy similar to the one in [4]. The concealment strategy is spatially causal, i.e., the decoder will only use the information from previously received packets in concealing a lost packet. When a packet is lost, the concealment motion vector for a macroblock (MB) in the lost packet is the median of the three motion vectors of its top-left, top, and top-right MBs. If the previous packet is also lost, then the concealment motion vector is zero, i.e., the MB in the same spatial location in the previously reconstructed frame is used to conceal the current loss.

4.3. End-to-End Distortion

Due to channel losses, the expected distortion can be calculated at the encoder as

$$E[D_k] = (1-\rho_k)E[D_{R,k}] + \rho_k E[D_{L,k}], \tag{4}$$

where $E[D_{R,k}]$ and $E[D_{L,k}]$ are the expected distortion when the k-th source packet is either received correctly or lost, respectively, and ρ_k is its loss probability. The relationship between the source packet loss probability, ρ, and transport packet loss probability, ϵ, depends on the specific transport packetization scheme chosen. Methods to calculate this can be found in [4, 10]. Note that both $D_{L,k}$ and $D_{R,k}$ are usually random variables. This is because, due to channel losses, the reference frames for inter-coding at the decoder and the encoder may not be the same. Also note that the calculation of $D_{L,k}$ depends on the specific error concealment strategy used at the decoder.

Assuming the mean squared error (MSE) criterion, the distortion measurement based on the ROPE (Recursive Optimal Per-pixel Estimate) algorithm [9] is used to recursively calculate the overall expected distortion level of each pixel. The image quality measure used is the peak signal-to-noise-ratio (PSNR), defined as PSNR$= 10\log\frac{255^2}{\text{MSE}}$ dB.

5. SOLUTION ALGORITHM

By using a Lagrange multiplier $\lambda \ge 0$, (3) can be converted into an unconstrained problem as,

$$\min_{\{\boldsymbol{\mu}\in\mathcal{Q},\boldsymbol{\nu}\in\mathcal{R}\}} \sum_{k=1}^{M} J_k^{(n)} = \sum_{k=1}^{M} E[D_k^{(n)}] + \lambda \sum_{k=1}^{M} T_k^{(n)}, \tag{5}$$

where M is the number of packets in the frame. Note that the transmission delay for the k-th packet, $T_k^{(n)} = B_k^{(n)}/R_T$, takes into account the associated channel bits used to protect this packet. The convex hull solution of this relaxed problem can be found by choosing an appropriate λ to satisfy the transmission delay constraint. This can be done using standard techniques such as a bisection search [12]. We can write the problem as:

$$\min_{\{\boldsymbol{\mu}\in\mathcal{Q},\boldsymbol{\nu}\in\mathcal{R}\}} \sum_{k=1}^{M} J_k^{(n)} = \min_{\{\boldsymbol{\nu}\in\mathcal{R}\}} \{ \min_{\{\boldsymbol{\mu}\in\mathcal{Q}\}} \sum_{k=1}^{M} J_k^{(n)}(\boldsymbol{\mu},\boldsymbol{\nu})\}, \quad (6)$$

where $J_k^{(n)} = E[D_k^{(n)}] + \lambda T_k^{(n)}$. Given a specific λ, the minimization of (6) can be divided into two steps: bit allocation for FEC and optimal mode selection for the current frame based on the remaining delay. Note that this differs from solving (1) and (2) in that the bit allocation for FEC takes into account the effect of this choice on source coding. The optimal mode selection can be found using a dynamic programming (DP) approach. The DP can be viewed as a shortest path problem in a trellis, where each stage corresponds to the mode selection for a given packet [11,12]. Note that by using the error concealment strategy described in Sect. 4.2, the distortion $E[D_k]$ depends on the encoding modes and probability of source packet loss selected for the previous source packet. Thus, the Lagrangian $\sum_{k=1}^{M} J_k^{(n)}(\boldsymbol{\mu},\boldsymbol{\nu})$ in (6) is not separable. In this case, the time complexity is $O(|M \times |\mathcal{R}| \times |\mathcal{Q}|^2)$, where $|\cdot|$ denotes the cardinality of the set inside [12].

6. EXPERIMENTAL RESULTS

In the simulations, we use an H.263+ codec to perform source coding, and we consider the QCIF format (176×144) Foreman sequence. Rate control is not implemented in the video streaming system. Thus, every frame has the same transmission delay constraint of one frame's duration, i.e., $T_0^{(n)} = T_F$. We assume that after 1 RTT, channel feedback is available to the encoder in the form of which packets are received or lost. We consider applications that require a short end-to-end delay, T_{max}, and the RTT is set equal to two frames. Under that situation, the feedback delay is long enough to preclude retransmissions.

Four systems are compared: i) system 1, which uses the proposed framework to jointly consider error resilient source coding and channel coding; ii) system 2, which performs error resilient source coding, but with fixed rate channel coding; iii) system 3, which performs only channel coding, but no error resilient source coding (i.e., source coding is not adapted to the modified channel characteristics after error recovery); and iv) system 4, which performs sequential JSCC. All four systems are optimized in the following manner: System 2 performs optimal error resilient source coding to adapt to the channel errors (with fixed rate channel coding). System 3 selects the optimal channel coding rate to perform FEC and does optimal source coding (without considering residue packet loss after channel coding) for a given bit budget. In the sequential JSCC, bit allocation between source and channel is performed with no awareness of error resilient source coding as in (1), and error resilient source coding is performed thereafter given the bit budget as in (2).

We illustrate the performance of the four systems in Fig. 1 at $R_T = 480$ kbps and frame rate $F = 15$ fps. Here, we plot the average PSNR against different packet loss rates. All four systems have the same transmission delay constraints and transmission rate. It can be seen in Fig. 1(a) that system 1 outperforms system 2 with different pre-selected channel coding rates. In addition, system 1 outperforms the optimized system 2 (the upper bound of system 2 with different pre-defined channel rates) with different channel coding rates by up to 0.3 dB. This is due to the flexibility of system 1 from varying the channel coding rate in response to the video content. As shown in Fig. 1(b), system 1 outperforms systems 3 and 4 with up to around 0.4 dB and 0.3 dB, respectively. The gain in system 1 compared to system 4 comes from the joint consideration of source coding and channel coding. The gain in system 4 in comparison to system 3 comes from the adaptation of source coding to the modified channel characteristics after error recovery (system 3 does not do error resilient source coding).

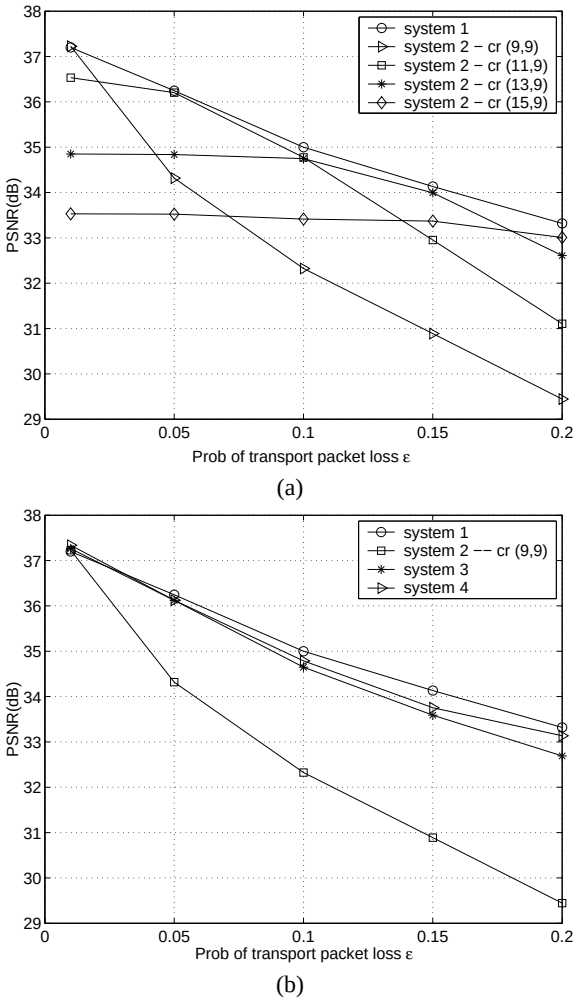

Fig. 1. Average PSNR vs. transport packet loss probability (a) System 1 vs. System 2 with indicated channel rates (b) System 1 vs. System 3 and 4 ($R_T = 480$ kbps, $F = 15$ fps, cr in the legend denotes channel rates).

In Fig. 2, we plot the PSNR against the transmission rate. It can be clearly seen that, as the transmission rate increases (i.e., the bit budget per frame increases), the gap between the performance of systems 1, 3 and 4 (with channel coding) and system 2 (without channel coding) also increases. This can be explained by the fact that the low bit budget restricts the ability to use channel coding, because a majority of the bits are needed for source coding. When

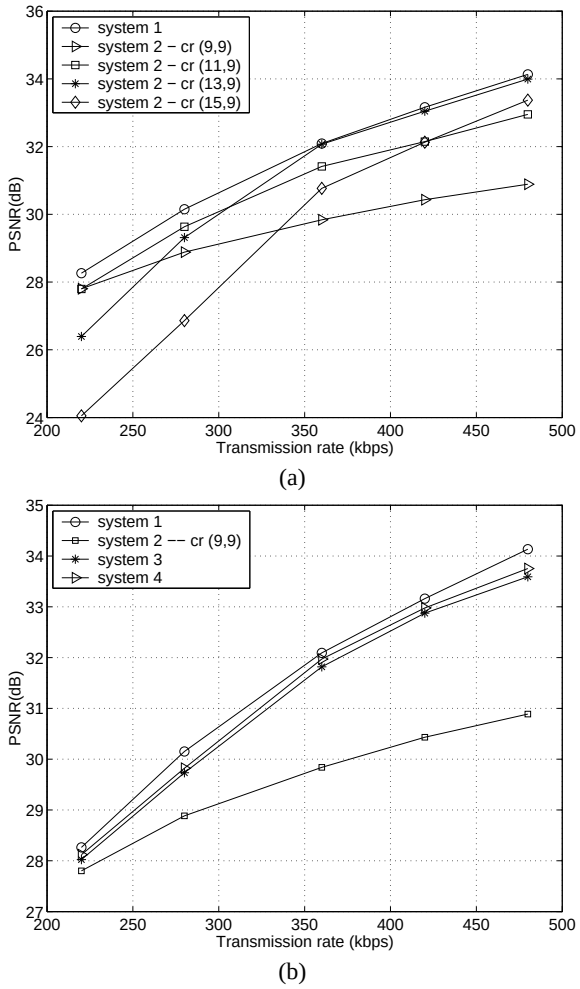

Fig. 2. Average PSNR vs. transmission rate (a) System 1 vs. System 2 with indicated channel rates (b) System 1 vs. System 3 and 4 ($\epsilon = 0.15$, $F = 15$ fps, cr in the legend denotes channel rates).

the bit budget gets larger, the system becomes more flexible in its ability to allocate bits to the channel in order to improve the overall performance. Again, as shown Fig. 2(a), system 1 outperforms system 2 with different pre-selected channel coding rates and also outperforms system 3 and 4 at various transmission rates.

Note that the gain of the IJSCC system (system 1) compared with system 3 (without performing error resilient source coding) or system 4 (the sequential JSCC) may not be very significant. This is because in all systems we perform the optimization by jointly considering several available error control components such as error concealment. Thus, absence of one of the error control components, as in system 3, or lack of the joint consideration of source and channel coding, as in system 4, may not have a very significant effect due to mitigation of other error control components in the system. In practical situations where computation resources are constrained, application of the integrated system may not be necessary if the gain over simpler techniques does not outweigh the additional computational complexity. In such cases, the value of the integrated system is that it provides an optimization benchmark against which the performances of other sub-optimal systems can be evaluated.

7. CONCLUSIONS

This paper presented an integrated joint source-channel coding framework, where error resilient source coding, channel coding, and error concealment are jointly considered in an integrated manner. We demonstrated through both analysis and simulations the advantages of the proposed IJSCC approach compared to a sequential JSCC approach. Although the gain may not be very significant, the optimal IJSCC framework serves as a useful tool in performance evaluation of sub-optimal systems.

8. REFERENCES

[1] J. Rosenberg and H. Schulzrinne, "An RTP payload format for generic forward error correction," Tech. Rep., Internet Engineering Task Force, Request for Comments (Proposed Standard) 2733, Dec. 1999.

[2] D. Wu, Y. T. Hou, and Y-Q Zhang, "Transporting real-time video over the Internet: Challenges and approaches," *Proc. IEEE*, vol. 88, pp. 1855–1877, Dec. 2000.

[3] T. Stockhammer and C. Buchner, "Progressive texture video streaming for lossy packet networks," in *Proc. International Packet Video Workshop*, Kyongju, Korea, April 2001.

[4] M. Gallant and F. Kossentini, "Rate-distortion optimized layered coding with unequal error protection for robust Internet video," *IEEE Trans. on Circ. and Syst. for Video Techn.*, vol. 11, no. 3, pp. 357–372, March 2001.

[5] 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. Circ. and Syst. for Video Techn.*, vol. 12, pp. 511–523, June 2002.

[6] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint source-channel coding for motion-compensated DCT-based SNR scalable video," *IEEE Trans on Image Processing*, vol. 11, pp. 1043–1052, Sept. 2002.

[7] G. Cheung and A. Zakhor, "Bit allocation for joint source/channel coding of scalable video," *IEEE Trans. Image Processing*, vol. 9, pp. 340–356, March 2000.

[8] J. Kim, R. M. Mersereau, and Y. Altunbasak, "Error-resilient image and video transmission over the Internet using unequal error protection," *IEEE Trans. Image Processing*, vol. 12, pp. 121–131, Feb. 2003.

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

[10] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Rate-distortion optimized hybrid error control for real-time packetized video transmission," in *Proc. IEEE Int. Conf. Communications (ICC'04)*, June 2004, accepted.

[11] F. Zhai, C. E. Luna, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and packet classification for real-time video transmission over differentiated services networks," *IEEE Trans. Multimedia*, 2004, accepted.

[12] G. M. Schuster and A. K. Katsaggelos, *Rate-Distortion Based Video Compression: Optimal Video Frame Compression and Object Boundary Encoding*, Kluwer Academic Publishers, 1997.