JOINT SOURCE-CHANNEL CODING FOR SCALABLE VIDEO USING MODELS OF RATE-DISTORTION FUNCTIONS

Lisimachos P. Kondi

Dept. of Electrical Engineering
University at Buffalo
The State University of New York
Buffalo, NY 14260

Aggelos K. Katsaggelos

Dept. of Electrical and Computer Engineering
Northwestern University
Evanston, IL 60208

ABSTRACT

A joint source-channel coding scheme for scalable video is developed in this paper. An SNR scalable video coder is used and Unequal Error Protection (UEP) is allowed for each scalable layer. Our problem is to allocate the available bit rate across scalable layers and, within each layer, between source and channel coding, while minimizing the end-to-end distortion of the received video sequence. The resulting optimization algorithm we propose utilizes universal rate-distortion characteristic plots. These plots show the contribution of each layer to the total distortion as a function of the source rate of the layer and the residual bit error rate (the error rate that remains after the use of channel coding). Models for these plots are proposed in order to reduce the computational complexity of the solution. Experimental results demonstrate the effectiveness of the proposed approach.

1. INTRODUCTION

During the past few years there has been an increasing interest in multimedia communications over different types of channels. In recent days a significant amount of research has been focused on multimedia transmission over wireless channels. This is a complex and challenging problem due to the multipath fading characteristics of the channel.

Source coding is concerned with the efficient representation of a signal. While bit errors in the uncompressed signal can cause minimal distortion, in its compressed format a single bit error can lead to significantly larger errors. Hence for transmission over an error prone channel, it is imperative that channel coding be employed to make the data more resilient to channel errors by increasing the redundancy.

Traditionally, source and channel coding have been considered independently. The reason behind this is Shannon's important information-theoretic result establishing the *principle of separability* [1]. It states that the design of source and channel coding can be separated without any loss in optimality as long as the source coding produces a bit rate that can be carried by the channel. This principle relies on the crucial assumption that the source and channel codes can be of arbitrary long lengths. In practical situations, due to limitations on the computational power and processing delays this assumption does not hold. It is then of benefit to consider the problem of source and channel coding jointly. Joint source-channel coding is an active research area. A review of joint source-channel coding for wireless channels can be found in [2].

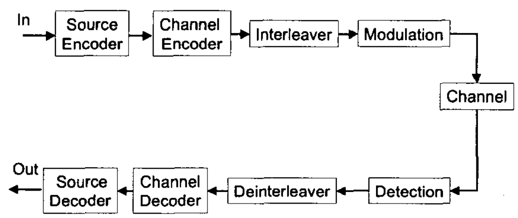

Fig. 1. Block diagram of a typical video transmission system.

In a compressed video bit stream the various parts of the bit stream are not equally important to the quality of the decoded video sequence. Thus, instead of protecting them equally, it would be advantageous to protect the most important parts of the bit stream more than the less important parts. This is the idea of data partitioning and Unequal Error Protection (UEP). In this work we apply UEP to the layers of a scalable bit stream.

The break-up of the bit stream into subsets of varying quality using a scalable codec lends itself naturally to employing an unequal error protection scheme. The base layer is typically better protected than the enhancement layers. This allows for added degrees of freedom in selecting the rates that will minimize the overall distortion. In [3], the benefits of using scalability in an error prone environment are shown by examining all the scalability modes supported by MPEG-2 in an ATM network.

The basic block structure of the system we are considering in this paper is shown in Fig. 1. We begin with a scalable video bit stream that is channel coded using a specified channel rate. This channel coded information is then interleaved and modulated for transmission over the channel. At the receiver the information is demodulated and deinterleaved. This received channel data is then decoded using a channel decoder and finally sent to the source decoder. In this paper we extend our previously reported results ([4, 5, 6]) by using a model for the Universal Rate Distortion Characteristics. The model used is similar to the one proposed in [7].

The rest of the paper is organized as follows. In section 2 we describe the elements of the video transmission system, i.e., scalable video coding (section 2.1), channel encoding (section 2.2), and wireless channel (section 2.3). In section 3, the joint source coding optimization algorithm is described. In section 4, experimental results are presented and in section 5, conclusions are drawn.

0-7803-7041-4/01/$10.00 ©2001 IEEE

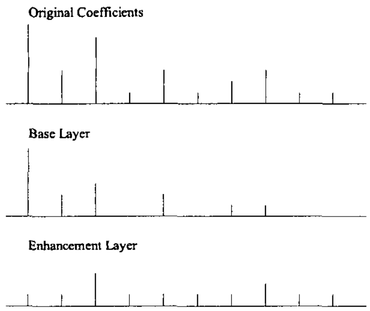

Fig. 2. Proposed partitioning of DCT coefficients for SNR scalability.

2. VIDEO TRANSMISSION SYSTEM

2.1. SCALABLE VIDEO CODING

A scalable video codec produces a bit stream which can be divided into embedded subsets, which can be independently decoded to provide video sequences of increasing quality. Thus, a single compression operation can produce bit streams with different rates and reconstructed quality. A subset of the original bit stream can be initially transmitted to provide a base layer quality with extra layers subsequently transmitted as enhancement layers.

There are three main types of scalability: Signal-to-Noise Ratio (SNR), spatial and temporal. In SNR scalability, the enhancement in quality translates in an increase in the SNR of the reconstructed video sequence, while in spatial and temporal scalability the spatial and temporal resolution, respectively, is increased. We describe next a method for SNR scalability which requires only a single DCT and quantization step [4, 8].

In the proposed SNR scalable video encoder, after motion compensation, the DCT of the prediction error is taken, as in a non-scalable encoder. Then, a value is subtracted from each quantization coefficient index. Subtracting a value from a coefficient reduces the number of bits required for its transmission but clearly increases the distortion. The number of bits required is decreased because for the Variable Length Code (VLC) tables proposed by the standards, the length of the codeword generally increases with the quantization level index. The coefficient indices are transmitted as base layer (along with the motion vectors and other overhead information) and the subtracted values are transmitted as enhancement layer. The decoder reconstructs the quantized DCT coefficients by adding the subtracted values (if the enhancement layer is available) to the values it received with the base layer. If more than two scalable layers are required, the values subtracted for the creation of the base layer are further broken into other values. For example, if we want to transmit a coefficient with magnitude of level 9 using three layers, we can transmit level 5 as base layer, level 2 as first enhancement layer and level 2 as second enhancement layer. We have developed an optimal algorithm for determining the partitioning of the DCT coefficients which is based on Lagrangian optimization and Dynamic Programming [4, 8].

2.2. Channel Coding

Rate-Compatible Punctured Convolutional (RCPC) codes for channel coding are used in this work. Punctured convolutional codes are families of channel codes that are obtained by puncturing the output of a "mother" convolutional code [9]. Puncturing is the process of removing, or deleting, bits from the output sequence in a predefined manner so that fewer bits are transmitted than in the original code leading to a higher coding rate. The idea of puncturing was extended to include the concept of rate compatibility [10]. Rate compatibility requires that a higher rate code be a subset of a lower rate code, or that lower protection codes be embedded into higher protection codes. This is accomplished by puncturing a "mother" code of rate $1/n$ to achieve higher rates (less protection).

2.3. Wireless Channels

Wireless, or mobile, channels differ from the traditional Additive White Gaussian Noise (AWGN) and wired computer networks in the types of errors they introduce, as well as, in the severity of these errors. A characteristic feature of wireless channels is *multipath fading*. It is the resulting degradation when multiple versions of a signal are received from different directions at different times. For the case in which no direct line-of-sight is available as in most urban areas we refer to this channel as a Rayleigh fading channel in which the received signal is corrupted by multiplicative noise having Rayleigh distribution.

Thus, given a transmitted signal $u(t)$ over a slowly fading AWGN channel using Binary Phase Shift Keying (BPSK) modulation, the received signal $r(t)$ over a signaling period can be represented as

$$r(t) = \alpha(t) \exp^{-j\phi(t)} u(t) + z(t), \quad 0 \leq t \leq T, \qquad (1)$$

where $z(t)$ is white gaussian noise, $\alpha(t)$ is the attenuation factor due to fading over the signaling period T and $\phi(t)$ is the phase shift of the received signal. For this signal the attenuation factor $\alpha(t)$ is a Rayleigh random process with the phase shift $\phi(t)$ being uniformly distributed over the interval $(-\pi, \pi)$.

In this paper, we assume that $\alpha(t)$ and $\phi(t)$ are constant over a signaling period. In the case of BPSK modulation over a fading channel, if the received signal's phase can be estimated from the signal for coherent demodulation, the received signal can be recovered with the use of a matched filter [11].

3. OPTIMAL BIT ALLOCATION BETWEEN SOURCE AND CHANNEL CODING

It has now become clear that channel coding is necessary in order to provide reliable visual communications over a wireless channel. Thus, the available bit budget should be shared between source and channel coding. However, it is not obvious how the bit allocation should be performed. In this section we propose a way of optimally allocating the available bits between source and channel coding.

The formal statement of the problem we are solving has as follows: Given an overall bit rate R_{budget}, we want to optimally allocate bits between the source and channel coding such as the overall distortion D_{s+c} is minimized, that is,

$$\min D_{s+c} \text{ subject to } R_{s+c} \leq R_{budget}, \qquad (2)$$

where R_{s+c} is the total bit rate used for source and channel coding for all layers and D_{s+c} is the resulting *expected* distortion which is due to both source coding errors and channel errors. The distortion that is caused by source coding is due to quantization and is deterministic. However, the distortion due to channel errors is stochastic. Thus, the total distortion is also stochastic and we use its expected value.

For L scalable layers, R_{s+c} is equal to $R_{s+c} = \sum_{l=1}^{L} R_{s+c,l}$, where $R_{s+c,l}$ is the bit rate used for source and channel coding for scalable layer l. It is equal to $R_{s+c,l} = \frac{R_{s,l}}{R_{c,l}}$, where $R_{s,l}$ and $R_{c,l}$ are the source and channel rates, respectively, for scalable layer l. It should be emphasized that $R_{s,l}$ is in bits/s and $R_{c,l}$ is a dimensionless number.

The problem is a discrete optimization problem, that is, $R_{s,l}$ and $R_{c,l}$ can only take values from discrete sets $\mathbf{R}_s^l$ and $\mathbf{R}_c^l$, respectively, i.e., $R_{s,l} \in \mathbf{R}_s^l, R_{c,l} \in \mathbf{R}_c^l$.

We will now utilize Lagrangian optimization to solve the problem of Eq. (2). In order to reduce the computational complexity of the solution, it is useful to write the overall distortion D_{s+c} as the sum of distortions per scalable layers:

$$D_{s+c} = \sum_{l=1}^{L} D_{s+c,l}. \tag{3}$$

In a subband-based scalable codec, it is straightforward to express the distortion as the sum of distortions per layer since each layer corresponds to different transform coefficients. However, in our scalable codec, we need to redefine distortion per layer as the *differential improvement* of including the layer in the reconstruction. Therefore, in the absence of channel errors, only the distortion for layer 1 (base layer) would be positive and the distortions for all other layers would be negative since inclusion of these layers reduces the Mean Squared Error (MSE).

Another observation that should be made is that the differential improvement in the MSE that the inclusion of a scalable layer causes depends on the rates of the previous layers. For example, for a two layer case, an enhancement layer of 28 kbps will cause a different improvement in the MSE depending on the rate used for the base layer. The differential improvement depends on how good the picture quality was to start with before the inclusion of the next scalable layer. Therefore, Eq. (3) can be written as $D_{s+c} = \sum_{l=1}^{L} D_{s+c,l}(R_{s+c,1}, \ldots, R_{s+c,l})$.

We are now ready to utilize Lagrangian optimization in order to solve the optimization problem. The constrained problem of Eq. (2) is transformed into the unconstrained problem of minimizing

$$J(\lambda) = D_{s+c} + \lambda R_{s+c}. \tag{4}$$

Our problem now reduces to finding the Operational Rate-Distortion Functions (ORDF)$D_{s+c,l}(., ..., .)$ for each scalable layer. Thus, we should experimentally obtain the expected distortion for each layer for all possible combinations of source and channel rates and all possible channel conditions. This would become prohibitively complex for even a small number of admissible source and channel rates and channel conditions. Thus, we have chosen to relax the optimality of the algorithm and utilize *Universal Rate-Distortion Characteristics* (URDC). These characteristics show the expected distortion per layer as a function of the bit error rate (after channel coding). In this paper,we assume the following model for the URDC.

$$D_{s+c,l} = a[\log_{10}(\frac{1}{P_e})]^b \tag{5}$$

where a and b are such that the square of the approximation error is minimized. Thus, instead of calculating the URDCs using experimental results for all possible bit error rates of interest, we experimentally calculate the expected distortion for a few bit error rates and use the model to approximate the distortion for other bit error rates. Assuming two scalable layers, three choices of source coding rates per layer and three chanel coding rates per layer, three URDCs would be required for the base layer, one for each source coding rate. Furthermore, nine URDCs would be required for the enhancement layer, since for each admissible source coding rate for the enhancement layer, we need three URDCs, one for each selection of base layer source coding rate. We need to specify a different parameter set (a, b) for each URDC.

We also need to obtain plots that show the channel bit error probability (bit error rate after channel coding) as a function of source coding rates and channel parameters. These should only depend on the channel model used and the channel codes. For the case of a Rayleigh fading channel assuming perfect interleaving, the channel parameter is the SNR per bit $\frac{E_b}{N_0}E[\alpha^2]$. We will call these plots *Channel Error Plots* (CEP). By combining the Channel Error Plots and the Universal Rate-Distortion Characteristics, we can estimate the Operational Rate-Distortion Functions. More details can be found in [4, 5, 6].

4. EXPERIMENTAL RESULTS

We performed the above optimization procedure using the proposed model for the Universal Rate-Distortion Curves. Three data points were used to obtain the parameters a and b for each ORDC. Those points corresponded to bit error rates of 10^{-7}, 10^{-6} and 10^{-5}. The data points were obtained using repeated experiments (30 runs) and taking the average distortion. Figure 3 shows the expected distortion versus the bit error rate for an enhancement layer source rate of 28 kbps assuming that the base layer source rate was also 28 kbps. Both the results of the model and the actual points used to obtain the parameters of the model are shown. Note that to obtain the enhancement layer distortion, only the enhancement layer is corrupted with errors, since we assume that the total distortion is the sum of the base and enhancement layer distortions. Also, to obtain the ORDFs used for the optimization, we need to subtract the base layer distortion since the ORDFs of the enhancement layers are defined as the differential improvement of receiving the enhancement layer in addition to the base layer. It can be seen from the plot that the fit of the model to the data is good.

We also tested the proposed model by comparing the results of the optimal allocation between source and channel coding obtained without the model with results obtained using the model. We assumed a Rayleigh fading channel with additive Gaussian noise and perfect interleaving (i.e., we assume that the samples of the Rayleigh random process α_k are independent identically distributed (i.i.d.)). The modulation scheme was Binary Phase Shift Keying (BPSK) with coherent detection. Rate-Compatible Punctured Convolutional codes (RCPC codes) were used for channel coding. One base and one enhancement layer was assumed. Table 1 shows the optimal operating points (rate allocations in the convex hull of the overall rate-distortion curve) for a 10 dB Rayleigh fading channel as presented in [4], without using the proposed model. Source and channel rate combinations not appearing in this table are not optimal, i.e., there are entries in this table with both lower total rate and lower distortion. We chose the sets of admissible source rates for both layers to be $\mathbf{R}_s^1 = \mathbf{R}_s^2 = \{28, 42, 56\}$ kbps. The set of admissible channel rates was chosen to be $\mathbf{R}_c^1 = \mathbf{R}_c^2 = \{1/2, 2/3, 4/5\}$. We used RCPC codes with mother code rate $1/2$

R_{s+c} (kbps)	$R_{s,1}$ (kbps)	$R_{c,1}$	$R_{s,2}$ (kbps)	$R_{c,2}$
70	28	4/5	28	4/5
77	28	2/3	28	4/5
84	28	2/3	28	2/3
98	28	1/2	28	2/3
112	28	1/2	28	1/2
168	56	1/2	28	1/2
224	56	1/2	56	1/2

Table 1. Optimal rate allocation for two layer single-pass SNR scalable video over a 10 dB Rayleigh fading channel (from [4]).

R_{s+c} (kbps)	$R_{s,1}$ (kbps)	$R_{c,1}$	$R_{s,2}$ (kbps)	$R_{c,2}$
70	28	4/5	28	4/5
77	28	2/3	28	4/5
84	28	2/3	28	2/3
98	28	1/2	28	2/3
112	28	1/2	28	1/2
119	42	2/3	28	1/2
140	42	1/2	28	1/2
224	56	1/2	56	1/2

Table 2. Optimal rate allocation for two layer single-pass SNR scalable video over a 10 dB Rayleigh fading channel (using model).

from [10].

Table 2 shows the optimal operating points for the same setup using the proposed model. It can be seen that the results are very similar, thus proving the validity of the model. The only difference is that with the use of the model, the 168 kbps total rate is not on the convex hull of the solution, thus it does not appear in Table 2, while total rates of 119 kbps and 140 kbps are on the convex hull of the solution.

5. CONCLUSIONS

In this paper we extended our previous results reported in [4, 5, 6] by utilizing a parametric model for the Universal Rate-Distortion Characteristics. Without the use of the model, we would have to conduct simulations to find the expected distortion for every bit error rate that corresponds to an admissible channel coding rate. With the use of the model, we have shown that it is possible to only conduct simulations for a small number of bit error rates and use the model to estimate the distortion for other bit error rates, thus reducing the computational complexity of the solution.

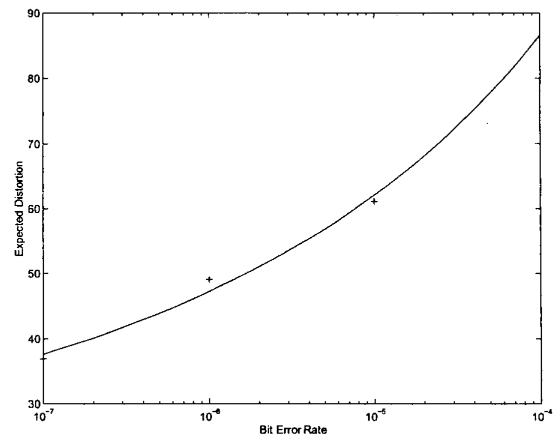

Fig. 3. Expected distortion versus bit error rate plot for an enhancement layer with a source rate of 28 kbps assuming a base layer with a source rate of 28 kbps. The result of the model is shown (solid line) along with the three points used to obtain the parameters of the model (crosses).

6. REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Interscience, 1991.

[2] K. Ramchandran and M. Vetterli, *Multiresolution Joint Source-Channel coding for Wireless Channels*, chapter in Wireless Communications: A Signal Processing Prespective, V.Poor and G. Wornell ed., Prentice-Hall, 1998.

[3] R. Aravind, M. R. Civanlar, and A. R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, no. 5, pp. 426–435, Oct. 1996.

[4] L. P. Kondi, *Low Bit Rate SNR Scalable Video Coding and Transmission*, Ph.D. thesis, Northwestern University, December 1999.

[5] F. Ishtiaq, L. P. Kondi, and A. K. Katsaggelos, "Joint source-channel coding for scalable video," Submitted to *IEEE Transactions on Image Processing*,19 99.

[6] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint source-channel coding for scalable video," in *Proceedings of the SPIE Conference on Image and Video Communications and Processing*,San Jose, CA, 2000, pp. 324–335.

[7] M. Bystrom and T.St ockhammer, "Modeling of operational distortion-rate characteristics for joint source-channel coding of video," in *Proceedings of the IEEE International Conference on Image Processing*,V ancouver, Canada, 2000.

[8] L. P.K ondi and A. K. Katsaggelos, "An optimal single pass SNR scalable video coder," in *Proceedings of the IEEE International Conference on Image Processing*, Kobe, Japan, 1999.

[9] J. B. Cain, G. C. Clark, and J. M. Geist, "Punctured convolutional codes of rate $(n-1)/n$ and simplified maximum likelihood decoding," *IEEE Transactions on Information Theory*, vol. IT-25, pp. 97–100, Jan. 1979.

[10] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Transactions on Communications*, vol. 36, no. 4, pp. 389–400, April 1988.

[11] T. S. Rappaport, *Wireless Communications*, Prentice-Hall, 1996.