JOINT SOURCE-CHANNEL CODING FOR SCALABLE VIDEO OVER DS-CDMA MULTIPATH FADING CHANNELS

Lisimachos P. Kondi, Stella N. Batalama*, Dimitris A. Pados*, and Aggelos K. Katsaggelos+*

* Dept. of Electrical Engineering
State University of New York at Buffalo
Buffalo, NY 14260
+ Dept. of Electrical and Computer Engineering
Northwestern University
Evanston, IL 60208

ABSTRACT

In this paper, we extend our previous work on joint source-channel coding to scalable video transmission over wireless direct-sequence code-division-multiple-access (DS-CDMA) multipath fading channels. An SNR scalable video coder is used and unequal error protection (UEP) is allowed for each scalable layer. At the reveiver-end an adaptive antenna array Auxiliary-Vector (AV) filter is utilized that provides space-time RAKE-type processing and multiple-access interference suppression. The choice of the AV receiver is dictated by realistic channel fading rates that limit the data record available for receiver adaptation and redesign. Our problem is to allocate the available bit rate of the user of interest between source and channel coding and across scalable layers, while minimizing the end-to-end distortion of the received video sequence. The optimization algorithm that we propose utilizes universal rate-distortion characteristic curves that 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 after channel coding). These plots can be approximated using appropriate functions to reduce the computational complexity of the solution.

1. INTRODUCTION

During the past few years there has been an increasing interest in multimedia communications over different types of channels. A significant amount of research has been focused on multimedia transmission over wireless channels. In this paper we consider joint source-channel coding for scalable video over wireless DS-CDMA multipath fading channels.

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 in error can lead to a sequence of errors upon decompression. Hence, for transmission over an error prone channel it is imperative that channel coding be employed to make the data resilient to channel errors by increasing the redundancy.

Traditionally, source and channel coding have been considered independently. The reason behind that 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 arbitrarily long lengths. Certainly, in practical situations 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].

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 importance 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 our previous work [3, 4], we assumed video transmission from one transmitter to one receiver using binary phase-shift keying (BPSK) modulation. The channel model was a non frequency selective Rayleigh fading channel. In this paper, we are assuming a direct-sequence code-division-multiple-access (DS-CDMA) system. Thus, the data are spread before BPSK modulation and multiple users transmit over the same frequency band. The channel model that we consider here is a frequency-selective (multipath) Rayleigh fading channel. At the receiver, we employ an adaptive antenna array Auxiliary-Vector (AV) linear filter that provides space-time RAKE-type processing (thus taking advantage of the multipath characteristics of the channel) and multiple-access interference suppression. The choice of the AV receiver is dictated by realistic channel fading rates that limit the data record available for receiver adaptation and redesign. Indeed, AV filter short-data-record estimators have been shown to exhibit superior bit-error-rate performance in comparison with LMS, RLS, or SMI adaptive filter implementations [5, 6, 7].

The basic block structure of the single-user video transmission system that we consider 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 spread by means of a spreading code and carrier-modulated for transmission over the channel. At the receiver the information is carrier-

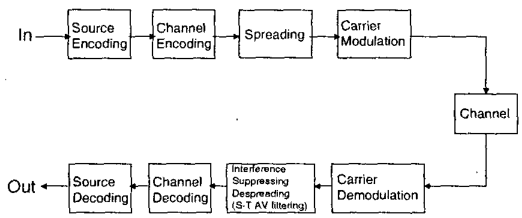

Fig. 1. Block diagram of the single user video transmission system.

demodulated and despread. These despread data are then decoded using a channel decoder and finally sent to the source decoder. The results of this work provide a realistic study of a complete end-to-end wireless video transmission link based on a realistic (multiuser) environment, realistic channel conditions/fading rates and state-of-the-art communication-link modules.

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), received signal (Section 2.3), and auxiliary-vector filtering (Section 2.4). In Section 3, the joint source coding optimization algorithm is described. In Section 4, experimental results are presented.

2. VIDEO TRANSMISSION SYSTEM

2.1. Scalable Video Coding

A scalable video codec produces a bit stream which can be divided into embedded subsets. The subsets 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. In this work, we utilize a method for SNR scalability which requires only a single discrete cosine transform (DCT) and quantization step. More details can be found in [8, 9].

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 [10]. 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 [11]. 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. Received Signal

The time-varying multipath characteristics of mobile radio communications suggest a channel model that undergoes a process known as fading. The baseband received signal at each antenna element m, $m = 1, \ldots, M$, is viewed as the aggregate of the multipath received SS signal of interest with signature code $\mathbf{S}_0$ of length L (if T is the symbol period and T_c is the chip period then $L = T/T_c$), $K-1$ multipath received DS-SS interferers with unknown signatures $\mathbf{S}_k$, $k = 1, \ldots, K-1$, and white Gaussian noise. For notational simplicity and without loss of generality, we choose a chip-synchronous signal set-up. We assume that the multipath spread is of the order of a few chip intervals, P, and since the signal is bandlimited to $B = 1/2T_c$ the lowpass channel can be represented as a tapped delay line with $P+1$ taps spaced at chip intervals T_c. After conventional chip-matched filtering and sampling at the chip rate over a multipath extended symbol interval of $L+P$ chips, the $L+P$ data samples from the mth antenna element, $m = 1, \ldots, M$, are organized in the form of a vector $\mathbf{r}_m$ given by $\mathbf{r}_m = \sum_{k=0}^{K-1} \sum_{p=0}^{P} c_{k,p} \sqrt{E_k} (b_k \mathbf{S}_{k,p} + b_k^- \mathbf{S}_{k,p}^- + b_k^+ \mathbf{S}_{k,p}^+) \mathbf{a}_{k,p}[\mathrm{m}] + \mathbf{n}$, $m = 1, \ldots, M$, where, with respect to the kth SS signal, E_k is the transmitted energy, b_k, b_k^-, and b_k^+ are the present, the previous, and the following transmitted bit, respectively, and $\{c_{k,p}\}$ are the coefficients of the frequency-selective slowly fading channel modeled as independent zero-mean complex Gaussian random variables that are assumed to remain constant over several symbol intervals. $\mathbf{S}_{k,p}$ represents the 0-padded by P, p-cyclic-shifted version of the signature of the kth SS signal $\mathbf{S}_k$, $\mathbf{S}_{k,p}^-$ is the 0-filled $(L-p)$-left-shifted version of $\mathbf{S}_{k,0}$, and $\mathbf{S}_{k,p}^+$ is the 0-filled $(L-p)$-right-shifted version of $\mathbf{S}_{k,0}$. Finally, $\mathbf{n}$ represents additive complex Gaussian noise and $\mathbf{a}_{k,p}[\mathrm{m}]$ is the mth coordinate of the kth SS signal, pth path, array response vector: $\mathbf{a}_{k,p}[\mathrm{m}] = e^{j2\pi(m-1)\frac{\sin\theta_{k,p} d}{\lambda}}$, $m = 1, \ldots, M$, where $\theta_{k,p}$ identifies the angle of arrival of the pth path of the kth SS signal, λ is the carrier wavelength, and d is the element spacing (usually $d = \lambda/2$).

To avoid in the sequel cumbersome 2-D data notation and filtering operations, we decide at this point to "vectorize" the $(L+P) \times M$ space-time data matrix $[\mathbf{r}_1 \ \mathbf{r}_2 \ \cdots \ \mathbf{r}_M]$ by sequencing all matrix columns in the form of a single $(L+P)M$-long column vector: $\mathbf{r}_{(\mathrm{L+P})\mathrm{M}\times 1} = Vec\{[\mathbf{r}_1, \ \mathbf{r}_2, \ \cdots, \ \mathbf{r}_M]_{(\mathrm{L+P})\times\mathrm{M}}\}$. From now on, $\mathbf{r}$ denotes the joint space-time data in the $\mathcal{C}^{(L+P)M}$ complex vector domain.

For conceptual and notational simplicity we may rewrite the vectorized space-time data equation as follows: $\mathbf{r} = \sqrt{E_0} b_0 \mathbf{w}_{\text{R-MF}} + \mathbf{I} + \mathbf{n}$ where $\mathbf{w}_{\text{R-MF}} = E_{b_0}\{\mathbf{r}\, b_0\} = Vec\{[\ \sum_{p=0}^{P} c_{0,p} \mathbf{S}_{0,p}, \mathbf{a}_{0,p}[1], \cdots, \cdots, \sum_{p=0}^{P} c_{0,p} \mathbf{S}_{0,p} \mathbf{a}_{0,p}[\mathrm{M}]\ ]\}$ is the effective space-time signature of the SS signal of interest (signal-*0*) and $\mathbf{I}$ identifies comprehensively both the Inter-Symbol and the SS interference present in $\mathbf{r}$ ($E_{b_0}\{\cdot\}$ denotes statistical expectation with respect to b_0). We use the subscript R-MF in our effective S-T signature notation to make a direct association with the RAKE Matched-Filter time-domain receiver that is known to correlate the signature $\mathbf{S}_0$ with $P+1$ size-L shifted windows of the received signal (that correspond to the $P+1$ paths of the channel), appropriately weighted by the conjugated channel coefficients $c_{0,p}, p = 0, \ldots, P$. In our notation, the generalized S-T RAKE operation corresponds to linear filtering of the form $\mathbf{w}_{\text{R-MF}}^H \mathbf{r}$, where H denotes the Hermitian operation.

2.4. Auxiliary-Vector Filtering

After carrier demodulation, chip-matched filtering, and chip-rate sampling, auxiliary-vector (AV) filtering (based on non-orthogonal AV components as in [7]) provides multiple-access-interference suppressing despreading. We recall that the AV algorithm generates a sequence of AV filters making use of two basic principles: (i) The maximum magnitude cross-correlation criterion for the evaluation of the auxiliary vectors and (ii) the conditional mean-square optimization criterion for the evaluation of the scalar AV weights. In summary, constrained to be distortionless in a vector direction of interest, a new element of the AV filter sequence is obtained as the sum of the previous filter in the sequence and a weighted auxiliary vector that is orthogonal to the constraint vector. The auxiliary-vector direction is chosen to maximize the magnitude of the statistical cross-correlation between the previous filter output and the projection of the data onto the auxiliary vector itself, while the corresponding scalar weight is chosen to minimize the new filter output variance. The sequence of filters obtained by the above conditional optimization procedure was shown to converge to the minimum-variance-distortionless-response (MVDR) solution under ideal setups (perfectly known input autocovariance matrix) [7]. When filter estimation based on a finite data record is performed by utilizing the sample-average estimate of the ideal input autocovariance matrix, the sequence of AV filter estimates was shown to converge to the sample-matrix-inversion (SMI) MVDR filter estimate [7]. Viewed as a sequence of estimators of the ideal MVDR filter, the sequence of AV-filter estimators exhibits the following characteristic that makes it a favorable choice for multiple-access-interference suppression in rapidly changing communication environments: The early non-asymptotic elements in the sequence offer favorable bias/variance balance and outperform significantly in mean-square filter estimation error the LMS, RLS or SMI adaptive filter implementations [7]. We conclude this section with a brief presentation of the AV algorithm. The AV filter sequence $\{\mathbf{w}_{AV}^{(d)}\}$, $d = 0, 1, 2, \ldots$ is initialized at the S-T vector direction of interest, i.e. the normalized S-T RAKE matched filter for the SS signal of interest, $\mathbf{w}_{AV}^0 = \mathbf{w}_{\text{R-MF}}$. Then, for $d = 1, 2, \ldots$, $\mathbf{w}_{AV}^{(d)} = \mathbf{w}_{AV}^{(d-1)} - \mu_d \mathbf{g}_d$ where $\mathbf{g}_d = (I - \mathbf{w}_{\text{R-MF}}\mathbf{w}_{\text{R-MF}}^H)\mathbf{A}\mathbf{w}_{AV}^{(d-1)}$, and $\mu_d = \frac{\mathbf{g}_d^H \mathbf{A}\mathbf{w}_{AV}^{(d-1)}}{\mathbf{g}_d^H \mathbf{A}\mathbf{g}_d}$ (I denotes the identity matrix and $\mathbf{A}$ is the input space-time autocovariance matrix, $\mathbf{A} = E\{\mathbf{r}\mathbf{r}^H\}$). For the selection of the best AV filter in the sequence (best number of auxiliary vectors d), please see [12].

3. OPTIMAL BIT ALLOCATION BETWEEN SOURCE AND CHANNEL CODING

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

$$\min D_{s+c} \text{ subject to } R_{s+c} \leq R_{budget} \qquad (1)$$

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 the 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 the 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: $R_{s,l}$ and $R_{c,l}$ can only take values from the discrete sets $\mathbf{R}_s^l$ and $\mathbf{R}_c^l$, respectively.

The optimization problem in (1) can be solved by Lagrangian optimization. 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: $D_{s+c} = \sum_{l=1}^{L} D_{s+c,l}$. 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 MSE due to a given layer 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 before the inclusion of the next scalable layer. Therefore, the distortion per layer is better expressed as

$$D_{s+c} = \sum_{l=1}^{L} D_{s+c,l}(R_{s+c,1}, \ldots, R_{s+c,l}).$$

Utilizing Lagrangian optimization, the constrained problem in (1) is transformed into the unconstrained problem of minimizing $J(\lambda) = D_{s+c} + \lambda R_{s+c}$. Thus, our problem reduces to finding the Operational Rate-Distortion Functions (ORDF)$D_{s+c,l}(\cdot, \ldots, \cdot)$ for each scalable layer. One way to proceed is to experimentally obtain the expected distortion for each layer for all possible combinations of source and channel rates and all possible channel conditions. However, this may become prohibitively complex for even a small number of admissible source and channel rates and channel conditions. Thus, we choose to utilize the *universal rate-distortion characteristics* (URDC) at the expense of a slight deviation from the optimal performance. These characteristics show the expected distortion per layer as a function of the bit error rate (after channel coding). More information on this issue can be found in [3, 4]. The URDC curves can be approximated using appropriate functions to

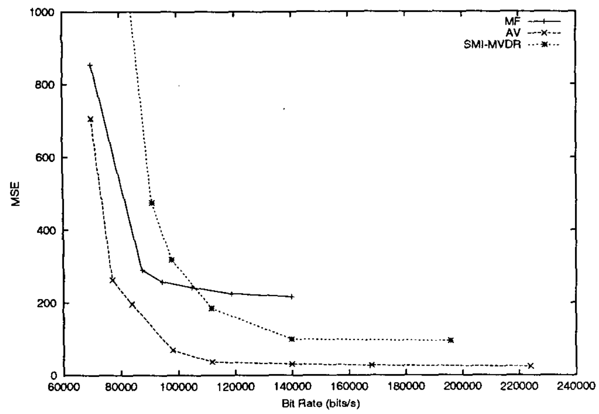

Fig. 2. Distortion versus transmitted bit rate for scalable video over DS-CDMA channels using three different receivers: Matched Filter (MF), Auxiliary Vector filter (AV) and the SMI-MVDR filter.

further reduce the computational complexity of the solution. In this work, we use the model $D_{s+c,l} = a[\log_{10}(\frac{1}{P_e})]^b$, where P_e is the bit error rate and a and b are found from experimental data such that the square of the approximation error is minimized [13].

4. EXPERIMENTAL RESULTS

For the implementation of the communications system of Fig. 1, we assumed 10 video users that employ direct-sequence spread-spectrum signaling (one is the user of interest and the other nine users are interferers). The SNR of the user of interest is fixed at 15 dB while the SNRs of the interferes are fixed at 14, 16, 14, 16, 14, 16, 15, 14 and 16 dB, respectively. The receiver utilizes an array of four antennas. We performed the optimization described in the previous section and utilized the model for the URDC curves from [13]. We assumed a two-layer SNR scalable video codec as described in Section 2.1. The admissible source coding rates for both layers were 28, 42 and 56 kbps. RCPC codes from [11] were used with rates of 1/2, 2/3 and 4/5. Three different receivers were assumed: The Matched Filter (MF), the conventional Sample-Matrix-Inversion Minimum-Variance-Distortionless-Response (SMI-MVDR) filter and the Auxiliary Vector (AV) filter with four AVs. For the optimization it is assumed that interleaving is performed such that the bit errors at the output of the channel decoder are independent. Figure 2 shows the overall distortion versus the total transmitted bit rate. We see that the receiver that employs AV-filtering outperforms the other two receivers for any given bit rate. Table 1 shows the optimal bit allocation for the case of the AV receiver.

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
140	28	1/2	42	1/2
168	28	1/2	56	1/2
224	56	1/2	56	1/2

Table 1. Optimal rate allocation for the receiver which employs AV-filtering.

5. REFERENCES

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

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

[3] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint source-channel coding for scalable video," Submitted to *IEEE Transactions on Image Processing*, 1999.

[4] 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.

[5] A. Kansal, S. N. Batalama, and D. A. Pados, "Adaptive maximum SINR RAKE filtering for DS-CDMA multipath fading channels," *IEEE J. Select. Areas Commmun.*, vol. 16, pp. 1765–1773, Dec. 1998, Special Issue on Signal Processing for Wireless Communications.

[6] D. A. Pados and S. N. Batalama, "Joint space-time auxiliary-vector filtering for DS-CDMA systems with antenna arrays," *IEEE Trans. Commmun.*, vol. 47, pp. 1406–1415, Sept. 1999.

[7] D. A. Pados and G. N. Karystinos, "An iterative algorithm for the computation of the MVDR filter," *IEEE Trans. Signal Proc.*, vol. 49, pp. 290–300, Feb. 2001.

[8] L. P. Kondi 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] L. P. Kondi and A. K. Katsaggelos, "An optimal single pass SNR scalable video coder," Submitted to *IEEE Transactions on Image Processing*, 1999.

[10] 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.

[11] 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.

[12] H. Qian and S. N. Batalama, "Data-record-based criteria for the selection of an auxiliary-vector estimator of the MVDR filter," in *Proceedings of the Thirty Fourth Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, 2000, pp. 802–807.

[13] L. P. Kondi and A. K. Katsaggelos, "Joint source-channel coding for scalable video using models of rate-distortion functions," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Salt Lake City, Utah, 2001.