Joint Source-Channel Coding for Motion-Compensated DCT-Based SNR Scalable Video

Lisimachos P. Kondi, *Member, IEEE*, Faisal Ishtiaq, and Aggelos K. Katsaggelos, *Fellow, IEEE*

Abstract—In this paper, we develop an approach toward joint source-channel coding for motion-compensated DCT-based scalable video coding and transmission. A framework for the optimal selection of the source and channel coding rates over all scalable layers is presented such that the overall distortion is minimized. The algorithm utilizes universal rate distortion characteristics which are obtained experimentally and show the sensitivity of the source encoder and decoder to channel errors. The proposed algorithm allocates the available bit rate between scalable layers and, within each layer, between source and channel coding. We present the results of this rate allocation algorithm for video transmission over a wireless channel using the H.263 Version 2 signal-to-noise ratio (SNR) scalable codec for source coding and rate-compatible punctured convolutional (RCPC) codes for channel coding. We discuss the performance of the algorithm with respect to the channel conditions, coding methodologies, layer rates, and number of layers.

Index Terms—Joint source-channel coding, scalable video, wireless video.

I. INTRODUCTION

DURING the past few years, there has been an increasing interest in multimedia communications over different types of channels, and in particular wireless channels [1]–[9]. This is a complex and challenging problem due to the multipath fading characteristics of the channel. Also, the distortion at the receiver is a function of both the lossy source coding technique and the errors introduced by the channel. Hence, it is not clear of how to best allocate a given bit budget between source and channel coding so that the resulting distortion at the receiver is minimized.

Source coding is concerned with the efficient representation of a signal. This is accomplished by reducing the redundancy in the signal as much as possible. While a bit error in the uncompressed signal can cause minimal distortion, in its compressed format a single bit error can lead to significantly larger errors. Although source coding is very effective in reducing the rate of the original sequence, it renders the compressed signal very sensitive to the impact of 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* [10]. 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. While being an important theoretical derivation, this principle relies on the crucial assumption that the source and channel codes can be of arbitrarily large 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 problems of source and channel coding jointly.

Joint video source-channel coding has been an active area of research. The focus of the techniques reported in the literature has been to minimize a given criterion (distortion, power consumption, delay, etc.) based on a given constraint. Among the many research directions, studies have focused on subband coding, scalable three-dimensional (3-D) subband coding, and cosine transformed video, [3], [4], [8], [9], [11]–[13]. Joint source-channel coding for high-definition television (HDTV) has been reported in [14]. While most techniques deal with the "digital" (source and channel coding in terms of bits) aspects of source-channel coding, "analog" aspects have also been investigated where modulation techniques are developed in conjunction with the source code and the channel in mind. Studies on embedded modulation techniques and embedded constellations are discussed in [15] and [16] among others. In [6], recent techniques on joint source-channel coding are surveyed and their integration into video compression standards is considered. In [7], the topic of video transmission over third-generation wireless communication systems is discussed. The authors use the wideband code division multiple access (W-CDMA) as a transport vehicle.

In a compressed video bitstream the various parts of the bitstream are not equally important. A natural way of protecting such a bitstream is by protecting more the bits that will impact the quality the most, such as frame headers, and less the

L. P. Kondi is with the Department of Electrical Engineering, State University of New York at Buffalo, Amherst, NY 14260 USA (e-mail: lkondi@eng.buffalo.edu).

F. Ishtiaq is with the Multimedia Communications Research Labs, Motorola Labs, Motorola, Inc., Schaumburg, IL 60196 USA (e-mail: faisal@labs.mot.com).

A. K. Katsaggelos is with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208 USA (e-mail: aggk@ece.northwestern.edu).

remaining bits, such as the DCT texture data. This concept of varying the protection of the bitstream according to its importance is called unequal error protection (UEP). In this work, we apply UEP to the layers of a scalable bitstream.

While scalability adds many functionalities to a bitstream, especially for heterogeneous networks, it is also well suited for enhanced error resilience. The breakup of a frame into subsets of varying quality lends itself naturally to employing an unequal error protection scheme, in which typically the base layer is better protected than the enhancement layers. This allows for added degrees of freedom in selecting the source and channel rates that will minimize the overall distortion. In [17], the benefits of scalability in an error prone environment are shown by examining all the scalability modes supported by MPEG-2 in an ATM network.

Joint source-channel coding of video involves many facets of communications, information theory, and signal processing. The basic block structure of the entire system with the individual elements involved is shown in Fig. 1. We begin with a scalable video bitstream 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. We will define each block in more detail as we progress through the paper in the development of an optimal joint source-channel coding algorithm for scalable video.

The paper is organized as follows. In Section II, we formulate the joint source-channel coding problem in detail. The rate allocation algorithm is presented in Section III. In Section IV we apply the proposed algorithm to the case of a wireless channel and rate-compatible punctured convolutional (RCPC) codes. In Section IV-A the main characteristics of the wireless channel are discussed while in Section IV-B the channel codes used are presented. In Section V, experimental results are presented. Finally, in Section VI conclusions are drawn.

II. PROBLEM FORMULATION

In this paper, we utilize joint source-channel coding for the effective selection of a source coding rate and a channel coding rate that will allow for the least amount of overall end-to-end distortion for a given total rate. The overall end-to-end distortion includes the distortion caused by both the source coding and the channel. In multiresolution and scalable video coding the problem has an added level of complexity in how to best allocate rates between the source and channel amongst the different layers. Toward finding an optimal solution to the source and channel rates for the layers, we formulate this problem into a bit, or resource, allocation problem with the objective that the overall distortion is minimized. We should note that while the distortion due to the source coding is deterministic in nature for a fixed set of quantizer choices, the distortion caused by the channel is of a stochastic nature. We therefore will use the term distortion to refer to the *expected* value of the distortion.
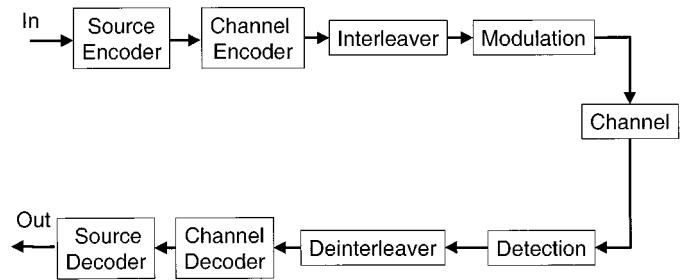

Fig. 1. Block diagram of a typical video transmission system.

A. Scalable Video Coding

A scalable video codec produces a bitstream 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 bitstreams with different rates and reconstructed quality. A small subset of the original bitstream can be initially transmitted to provide a base layer quality with extra layers subsequently transmitted as enhancement layers.

The approach presented in this paper is applicable to different signal-to-noise ratio (SNR) scalable coders, either supported by the standards or outside the standards. In the actual implementation of the system of Fig. 1, however, the H.263 video compression standard will be used for providing the SNR scalable bitstream (annex O of the standard).

SNR scalability in H.263 enables the transmission of the coding error (difference between the original and reconstructed picture) as an enhancement to the decoded picture, thus increasing the SNR of the decoded picture. There can be any number of enhancement layers. This is achieved by continuing to encode the difference between the original picture and the reconstructed picture including all prior enhancement layers.

If prediction is only formed from the lower layer, then the enhancement layer picture is called an EI-picture (Fig. 2). However, it is possible to create a modified bi-directionally predicted picture using both a prior enhancement layer picture (EI) and the current lower layer reference picture (P). These pictures are called "Enhancement" P-pictures, or EP-pictures. Both EI- and EP-pictures use no motion vectors for the prediction from the lower layer. However, as with normal P-pictures, EP-pictures use motion vectors when predicting from their temporally prior reference picture in the same layer (EI). Thus, new motion vectors have to be transmitted with the EP-picture.

B. Optimizing Problem

The problem we address is as follows. Given an overall coding rate, R_{s+c}, for both source and channel encoding over all scalable layers, we want to optimally allocate bits such that the total distortion D_{s+c} is minimized, that is,

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

For L layers, R_{s+c} is defined as

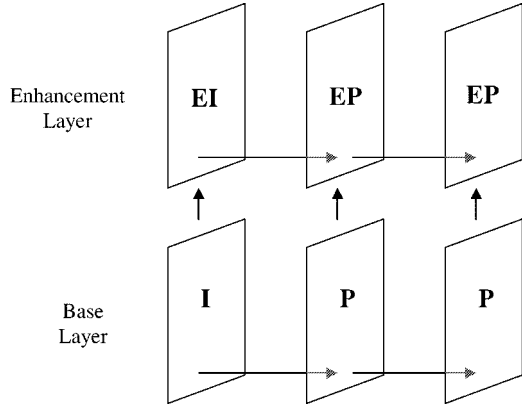$$R_{s+c} = \sum_{l=1}^{L} R_{s+c,l} \qquad (2)$$

Fig. 2. SNR scalability in H.263.

where $R_{s+c,\,l}$ is the combined source and channel rate for layer l. It is defined by

$$R_{s+c,\,l} = \frac{R_{s,l}}{R_{c,l}} \qquad (3)$$

with $R_{s,\,l}$ and $R_{c,\,l}$ being the individual source and channel rates for layer l, respectively, from the set of admissible rates for the given layer, R_s^l and R_c^l

$$R_{s,\,l} \in R_s^l = \{R_{s1}^l, R_{s2}^l, \ldots, R_{sM_l}^l\}$$
$$R_{c,\,l} \in R_c^l = \{R_{c1}^l, R_{c2}^l, \ldots, R_{cN_l}^l\} \qquad (4)$$

where M_l and N_l are the cardinalities of R_s^l and R_c^l, respectively. The source rates $R_{s,\,l}$ are in bits per second (bps) while the channel rates $R_{c,\,l}$ are the ratio of the number of information bits over the total number of bits, and they are therefore dimensionless numbers smaller than unity. The objective therefore of the optimization in (1) is to find the rates $R_{s,\,l}$ and $R_{c,\,l}$, for $l = 1, \ldots, L$, so that the overall rate does not exceed R_{budget}, while the overall distortion is minimized. In our implementation, the same protection is used for the whole scalable layer, i.e., no unequal error protection is applied within a layer. In a practical video compression system, the headers are more important than the motion vectors which, in turn, are more important than the DCT coefficients. A bit error in the headers will degrade the received video quality more than an error in the motion vectors. Also, an error in the motion vectors will cause more damage than an error in the DCT coefficients. Thus, we could assume, for example, three priority classes within each scalable layer: headers, motion vectors and DCT coefficients. Thus, it would be possible to incorporate unequal error protection within each layer by making $R_{c,\,l}$ a vector consisting of three channel coding rates, one for each priority class. This, however, would increase the computational complexity of the solution.

SNR scalability is supported by the H.263 [18] and MPEG-2 [19] video coding standards, as well as, other computationally efficient techniques [20]–[22]. 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, for a motion-compensated DCT-based codec, it is not as straightforward of how to define the distortion of each layer. In this paper, we define the

distortion per layer as the *differential improvement* due to the inclusion of this layer in the reconstruction. Therefore, in the absence of channel errors, only the distortion for layer one would be positive and the distortions for all other layers would be negative since inclusion of these layers reduces the mean squared error (MSE). Of course, in the presence of channel errors, it is possible for the distortion of any layer to be positive since inclusion of a badly damaged enhancement layer can increase the MSE.

Another observation is that the differential improvement in MSE that the inclusion of an enhancement layer causes depends not only on the rate of the layer but also on the rates of the previous layers. In other words, this differential improvement also depends on how good the picture quality was to start with before the inclusion of the current layer. Therefore, due to this dependency, the distortion for layer l is expressed as $D_{s+c,\,l}(R_{s+c,\,l}|R_{s,\,l-1}, \ldots, R_{s,\,1})$ and the overall distortion as

$$D_{s+c} = \sum_{l=1}^{L} D_{s+c,\,l}(R_{s+c,\,l}|R_{s,\,l-1}, \ldots, R_{s,\,1}) \qquad (5)$$

where

$$D_{s+c,\,l}(R_{s+c,\,l}|R_{s,\,l-1}, \ldots, R_{s,\,1})$$
$$= \tilde{D}_{s+c,\,l}(R_{s+c,\,l}|R_{s,\,l-1}, \ldots, R_{s,\,1})$$
$$- D_l^{uncorrupted}(R_{s,\,1}, \ldots, R_{s,\,l-1}) \qquad (6)$$

where $\tilde{D}_{s+c,\,l}(.)$ is the total distortion for layer l and $D_l^{uncorrupted}(.)$ is the distortion calculated when the first $l-1$ scalable layers are decoded without any bit errors. Thus, $D_{s+c,\,l}(.)$ denotes the differential improvement due to the inclusion of layer l in the reconstruction, as mentioned earlier.

III. OPTIMAL SCALABLE SOURCE-CHANNEL CODING WITH DEPENDENT LAYERS

The solution, D_{s+c}^*, to the constrained optimization problem of (1), (2), and (5) is obtained by converting it into an unconstrained one with the use of the Lagrangian functionals

$$J_1 = D_{s+c,\,1}^*(R_{s+c,\,1}) + \lambda R_{s+c,\,1}$$
$$\vdots$$
$$J_L = D_{s+c,\,L}^*(R_{s+c,\,L}|R_{s,\,L-1}, \ldots, R_{s,\,1}) + \lambda R_{s+c,\,L}. \qquad (7)$$

The unconstrained problem can now be solved using these L Lagrangian functionals, with the optimal solution being the argument of the following unconstrained minimization problem:

$$\min_{R_{s,l}, R_{c,l}, l=1, \ldots, L} [J_1 + J_2 + \cdots + J_L],$$

$$\text{where } R_{s+c,\,l} = \frac{R_{s,l}}{R_{c,l}}. \qquad (8)$$

The solution to this problem, R_{s+c}^* and hence D_{s+c}^*, is also the solution to the constrained problem of (1) if and only if $R_{s+c}^* = R_{budget}$ [23]. In practice, since there is only a finite set of choices for source and channel rates, it is not always possible to exactly meet R_{budget}. In this case, the solution is the bit

rate that is closest to R_{budget} while being lower than R_{budget}. By varying λ from zero to ∞, the result of (8) will trace out the operational rate-distortion curve for the system.

For each combination of available source rates at all previous layers, a set of operating points are obtained at the current layer l. The convex hull of these operating points is denoted by $D^*_{s+c,l}(.)$, that is,

$$D^*_{s+c,l}(R_{s+c,l}|R_{s,l-1}, \ldots, R_{s,1})$$
$$= C\{D_{s+c,l}(R_{s+c,l}|R_{s,l-1}, \ldots, R_{s,1})\} \quad (9)$$

where C represents the convex hull of the operational rate-distortion function (ORDF). A simple example is demonstrated by Fig. 3(a) and (b) for a two-layer case.

Assume that $D^*_{s+c,l}(.)$ is given for all L layers. We then use the Lagrangian approach for dependent layer distortion of (7) and (8). The arguments of the minimization in (8) give us the optimal bit allocation between scalable layers and between source and channel coding. Fig. 4 shows the overall ORDF that is obtained using the minimization algorithm.

A. Universal Rate-Distortion Characteristics

So far, we have formulated the problem and outlined the solution for the case of dependent layer distortions assuming that the optimal rate-distortion characteristics of the individual layers, $D^*_{s+c,l}(.)$, are given. We now describe the technique used to obtain the $D^*_{s+c,l}(.)$, $l = 1, \ldots, L$.

While it is possible to simulate transmission of the actual source coded data over a channel, gather statistics, and develop an optimal strategy based on these results, in practice, this leads to extremely high computational complexity and makes the process impractical in many ways. For every bitstream, we will have to simulate transmission of the data over the channel using all combinations of source and channel coding rates per scalable layer and for channel conditions of interest. Clearly, the computational complexity of this approach is prohibitive. To circumvent this problem, *universal* rate-distortion characteristics (URDCs) of the source coding scheme are utilized as proposed in [4] and [9]. This approach is described next.

Given a set of parameters for the channel, channel coding, and the particular modulation scheme, the probability of bit error, P_b, is calculated for the set of channel rates of interest. This can be done using simulations or by theoretical means. It establishes a reference as to the performance of the channel coding over the particular channel with the given parameters. Furthermore this channel performance analysis needs to be done only once. An example plot showing the performance of the channel coding as a function of a given channel parameter is shown in Fig. 5 for a set of channel coding rates, R_c. We will call these plots *channel characteristic plots*.

Toward calculating the impact of the errors due to both source coding and channel transmission on a set of data, it is realized that for a given set of preceding layer source rates, the distortion for a particular layer, $D_{s+c,l}$, given a particular source coding rate, $R_{s,l}$, is a function of the bit error rate. Thus the rate-distortion function of the layer, (given the preceding layer source rates) for a fixed source rate, $R_{s,l}$, is a function of the bit error rate (after channel decoding), P_b. It is then possible to
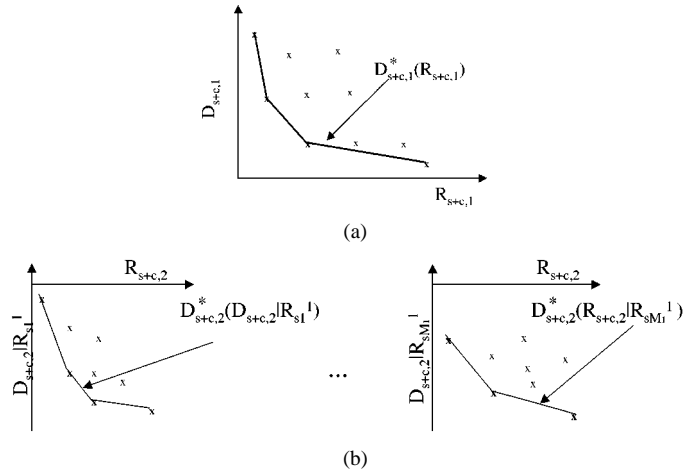

(a)

(b)

Fig. 3.   Rate-distortion functions of the (a) base and (b) enhancement layers for a two-layer example.
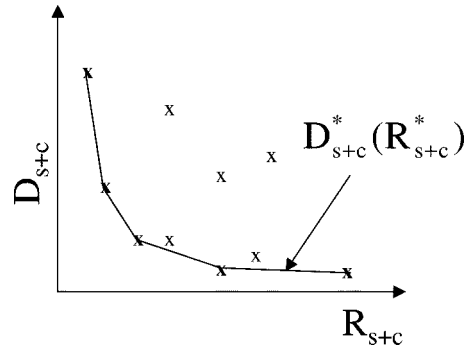

Fig. 4.   Rate-distortion profile of a given system. The convex hull of the points provides the optimal rate-distortion points.
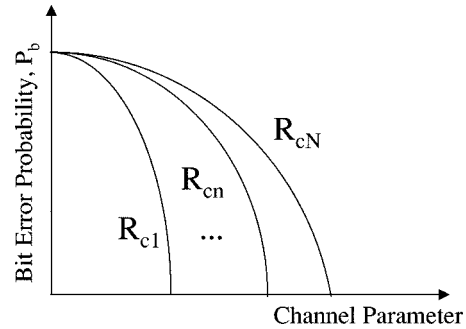

Fig. 5.   Typical channel characteristic plot, i.e., channel parameter versus bit error rate response characterizing the channel for a set of channel coding rates R_{cn} with $n = 1, \ldots, N$.

plot a family of $D_{s+c,l}$ versus $1/P_b$ curves given a set of source coding rates of interest as shown in Fig. 6. These are defined as the URDCs of the source. Due to the use of variable length codes in the video coding standards, it would be extremely difficult if not impossible to analytically obtain the URDCs. Thus, the URDCs are obtained experimentally using simulations. To obtain the URDC for $D_{s+c,l}$, layer l of the bitstream is corrupted with independent errors with bit error rate P_b. Layers $1, \ldots, l-1$ are not corrupted. The bitstream is then decoded and the mean squared error is computed. The experiment is repeated many times (in our case, 30 times). If $l > 1$, i.e., we are calculating the URDC for an enhancement layer, we need to

subtract the distortion of the first $l-1$ uncorrupted layers, since $D_{s+c,l}$ in this case is the differential improvement of including layer l, as mentioned earlier [see (6)].

Using the channel characteristic plots (Fig. 5) and the universal rate-distortion characteristics (Fig. 6), operational rate-distortion functions for each scalable layer are constructed as follows. First, for the given channel parameters, we use the channel characteristic plot to determine the resulting bit error rates for each of the available channel coding rates. Then, for each of these probability values, we use the URDCs to obtain the resulting distortion for each available source coding rates. By also obtaining the total rate R_{s+c} for each combination of source and channel codes, we have the rate-distortion operating points for the given channel conditions. The operational rate-distortion function is found by determining the convex hull of the operating points. It should be emphasized that the optimality is conditional on the universal rate-distortion characteristics and the assumption that the distortion for a particular layer is a function of the bit error rate. The channel parameters are the parameters that are necessary to characterize the channel and are different for each channel and modulation technique.

We now describe the proposed algorithm in more detail. We start by assuming that the state of the channel is given, determined by the channel parameters. Then, using the channel characteristic plot, for the set of channel rates of interest for layer l, $R_c^l = \{R_{c1}^l, R_{c2}^l, \ldots, R_{cN_l}^l\}$ the corresponding bit error rates $P_{ber}^l = \{P_{b1}^l, P_{b2}^l, \ldots, P_{bN_l}^l\}$ are acquired, where N_l is the number of admissible channel rates for layer l. Next, for each P_{bn}^l the distortions due to the source and channel for a set of source rates $R_s^l = \{R_{s1}^l, R_{s2}^l, \ldots, R_{sM_l}^l\}$ of interest for a particular combination of previous layer source rates are calculated, where M_l is the number of source rates of interest for layer l. As such, we now have the total rate and distortion points necessary for generating the convex hull for the layer conditioned upon a single combination of rates from the previous layers. Once the convex hull for all of the layers over all combinations of layer source rates have been generated, the minimization algorithm of (6) and (7) can be used to obtain the overall optimal rate-distortion breakdown.

IV. EXAMPLE: JOINT SOURCE-CHANNEL CODING FOR A WIRELESS CHANNEL USING RCPC CODES

A. Wireless Channels

In this section, we present essential elements on wireless channels for the further development and implementation of the scheme in Fig. 1. 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. Multipath fading occurs due to a number of factors of which some important ones are the presence and motion of objects reflecting the transmitted signal, the speed and motion of the receiver through this medium, and the bandwidth of the channel. Statistically, a wireless channel in which direct line-of-sight is available is
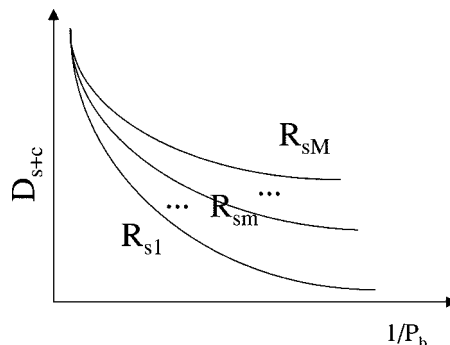

Fig. 6. Universal rate-distortion characteristics of the source; typical end-to-end distortion versus bit error rate response characterizing the source and channel for a set of source coding rates, R_{sm} for $m = 1, \ldots, M$.

referred to as a Rician fading channel where the distribution of the received signal follows a Rician probability density function. For the case in which no direct line-of-sight is available as in most urban areas the channel is referred to as a Rayleigh fading channel in which the distribution of the received signal follows a Rayleigh distribution. Its probability density function (pdf) is given by

$$p(r) = \begin{cases} \dfrac{r}{\sigma^2} \exp\left(-\dfrac{r^2}{2\sigma^2}\right), & (0 \leq r \leq \infty) \\ 0, & (r \leq 0) \end{cases} \quad (10)$$

where σ^2 is the time averaged power of the signal before detection. In the work presented in this paper, a Rayleigh fading channel is assumed.

Due to multiple reflections of the transmitted signal and the delay incurred with each reflected signal, the received signal is attenuated and delayed. Thus, given a transmitted signal $u(t)$ over a slowly fading channel with additive white Gaussian noise using binary phase shift keying (BPSK) modulation, the received signal $r(t)$ over a signaling period can be represented as

$$r(t) = \alpha(t) \exp^{-j\phi(t)} u(t) + z(t), \qquad 0 \leq t \leq T \quad (11)$$

where $z(t)$ is the complex valued 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 distributed random process with the phase shift $\phi(t)$ being uniformly distributed over the interval $(-\pi, \pi)$. For a slow fading Rayleigh channel, we can assume that $\alpha(t)$ and $\phi(t)$ are constant over a signaling interval.

In the case of binary phase shift keying modulation (BPSK) 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 [24].

B. Channel Coding

Rate-compatible punctured convolutional (RCPC) codes for channel coding are used in this work. Convolutional coding is accomplished by convolving the source data with a convolutional matrix, G. In essence rather than having a number of channel coded symbols for the corresponding block of source

TABLE I
PERFORMANCE OF RCPC OVER A RAYLEIGH
FADING CHANNEL WITH INTERLEAVING

Channel Rate R_c	1/2	2/3	4/5
Bit Error Probability P_b 10 dB	3.2×10^{-6}	9.1×10^{-5}	1.03×10^{-2}
Bit Error Probability P_b 20 dB	5.0×10^{-10}	8×10^{-7}	8.4×10^{-6}
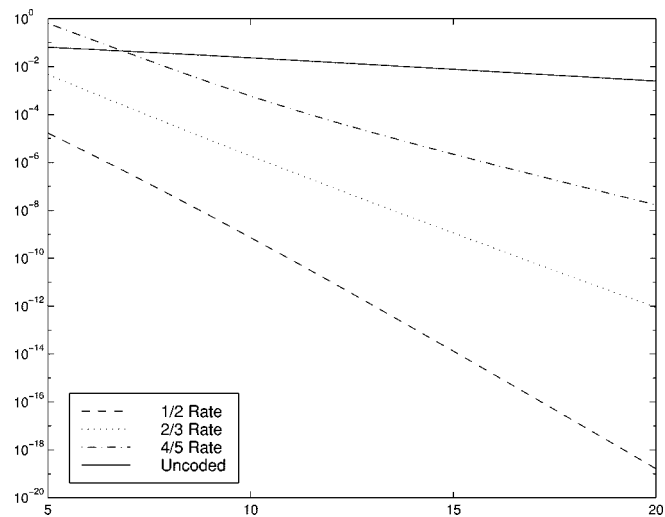
Fig. 7.   Theoretical bit error rates for RCPC codes using BPSK modulation over a Rayleigh fading channel.

Fig. 8.   Coded bit error rate versus uncoded bit error rate for RCPC codes using BPSK modulation over a Rayleigh fading channel.

Fig. 9.   Comparison of optimal rate-distortion characteristics of H.263 SNR scalability over a 10-dB Rayleigh channel using both low enhancement rates and high enhancement rates.

We begin by testing the algorithm with a 10 dB channel. The performance of 2 and 3 layer scalability at both the low enhancement rates and at high enhancement rates, as well as, the performance of a nonscalable bitstream are tested. The results are shown in Fig. 9. In comparing two- and three-layer scalability, it can be seen that at lower bitrates, two-layer scalability performs better for both the low and high enhancement rates. However, as the bitrates increase, the three-layer begins to perform better. It can also be seen that at bitrates above 120 Kbps, the three-layer, low enhancement rate, outperforms the two-layer high enhancement rates. At higher rates, the low and the high enhancement rates perform almost identically while at low bitrates the low enhancement rates clearly outperform the high enhancement rates. In comparing the scalable results to the nonscalable results, we note that for bitrates below 112 Kbps, scalability clearly does better in handling errors. As the rates increase the performance of scalable coding is comparable to, if not better than, nonscalable coding.

At 20 dB the subtleties of the rate allocation algorithm become more pronounced as seen in Fig. 10. Here, it is quite obvious that at these bitrates, the use of scalability can be effectively justified. Except for the 88 Kbps case, the scalable results are better than the nonscalable results. In comparing two versus three layers, the three-layer codec outperforms the two-layer codec at higher bitrates while the two-layer codec results in better distortion at lower bitrates. An interesting aspect of the algorithm is that for both the two- and three-layer codecs, the low enhancement rates give better results at lower rates and as the rate increases, there is a definite switch-over point beyond which the high rates perform better. This result is also present at 10 dB but less apparent.
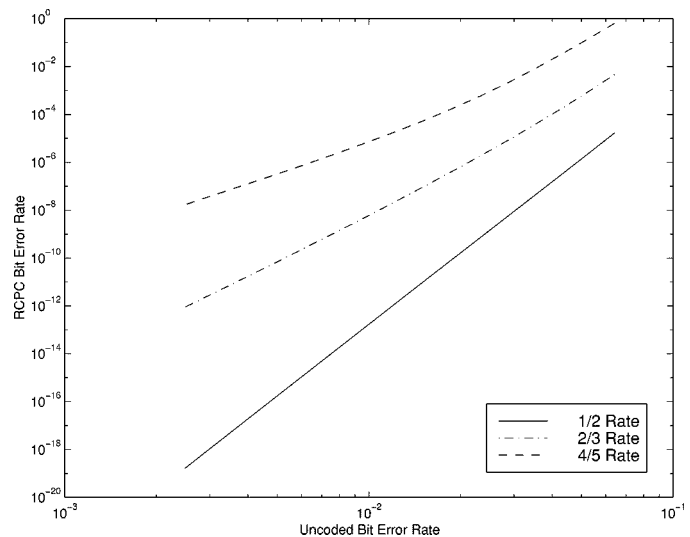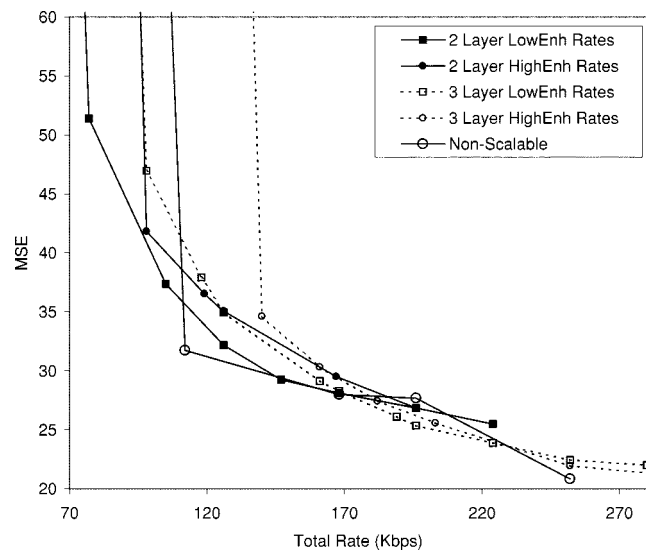
The optimal rate selections for the individual layers and type of coding are detailed in Tables II and III for the three-layer low enhancement rates case at 20 dB and 10 dB, respectively. One aspect that stands out is the fact that the distortions due to the channel coding dominate. This is especially apparent with the 10 dB channel where the 4/5 rate code is not enough to sustain any reasonable level of error protection. In this situation, for the three-layer case, all layers must be brought out of the 4/5 rate before any improvements can be seen. It is easier for the two layers since this can be done at a lower total rate. However, as the rates increase the performance of the layers converge and actually cross as the three layers are able to come up to a higher channel coding rate. At 20 dB the channel does not require a high level of protection and we see that the difference between the two and three layers is comparatively less at the lower rates. It is clear that the lowest channel rate possible is first selected for the base layer and as more overall rate is available, the emphasis
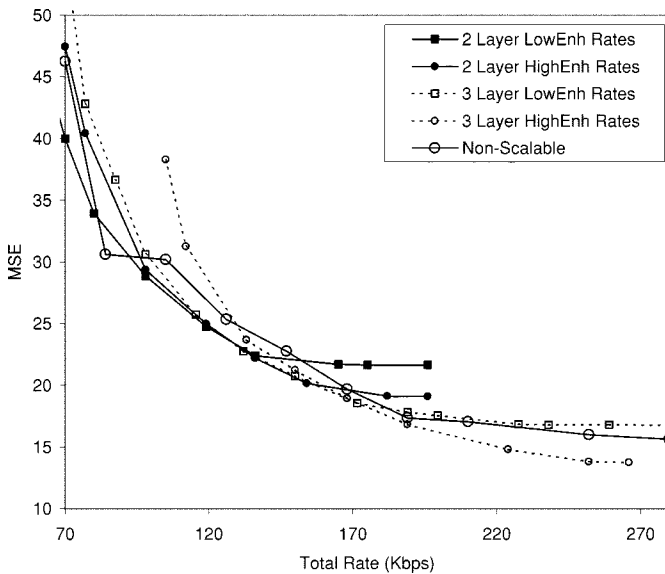
Fig. 10. Comparison of optimal rate-distortion characteristics of H.263 SNR scalability over a 20 dB Rayleigh channel using both low enhancement rates and high enhancement rates.

TABLE II
OPTIMAL RATE BREAKDOWN OF THREE-LAYER LOW ENHANCEMENT RATE H.263 SCALABLE VIDEO OVER A 20-dB RAYLEIGH FADING CHANNEL

R_{s+c}^*	D_{s+c}^*	$R_{s,1}$	$R_{c,1}$	$R_{s,2}$	$R_{c,2}$	$R_{s,3}$	$R_{c,3}$
70	54.96	28	4/5	14	4/5	14	4/5
88	36.65	42	4/5	14	4/5	14	4/5
116	25.73	42	2/3	14	4/5	28	4/5
132	22.77	42	2/3	14	4/5	42	4/5
150	20.74	42	2/3	28	4/5	42	4/5
189	17.81	56	2/3	42	4/5	42	4/5
228	16.85	56	1/2	42	4/5	42	2/3

TABLE III
OPTIMAL RATE BREAKDOWN OF THREE-LAYER LOW ENHANCEMENT RATE H.263 SCALABLE VIDEO OVER A 10-dB RAYLEIGH FADING CHANNEL

| R_{s+c}^*	D_{s+c}^*	$R_{s,1}$	$R_{c,1}$	$R_{s,2}$	$R_{c,2}$	$R_{s,3}$	$R_{c,3}$
70	3178	28	4/5	14	4/5	14	4/5
98	46.96	28	1/2	14	2/3	14	2/3
118	37.91	28	1/2	14	2/3	28	2/3
126	34.95	42	1/2	14	2/3	14	2/3
161	29.12	42	1/2	14	2/3	28	1/2
189	26.10	42	1/2	28	2/3	42	2/3
224	23.88	42	1/2	42	1/2	28	1/2

TABLE IV
COMPARISON OF ACTUAL TRANSMISSION VERSUS THE DISTORTION OF THE UNCORRUPTED SEQUENCE AND THE DISTORTION PROVIDED BY THE OPTIMAL RATE ALLOCATION ALGORITHM FOR THE "FOREMAN" SEQUENCE OVER A 10 dB AND 20 dB RAYLEIGH FADING CHANNEL

10 dB Channel	Distortion (MSE)		
	Uncorrupted	Optimization	Actual
H.263 Codec (126 Kbps)	28.80	35.05	33.65
20 dB Channel	Distortion (MSE)		
	Uncorrupted	Optimization	Actual
H.263 Codec (119 Kbps)	24.01	24.97	24.55

(a)

(b)

Fig. 11. Decoded H.263 frame 46 of the "Foreman" sequence (a) without any channel errors and (b) 10 dB Rayleigh fading channel. The base layer is 42 Kbps protected with a channel rate of 1/2 and the enhancement layer is 28 Kbps with 2/3 rate channel protection.

on the enhancement layer increases. However, in all cases, the emphasis rests clearly on the base layer. Intuitively, this is the obvious mode of rate selection. After all, the base layer has the ultimate decision on the quality of the overall video. Regardless of how well an enhancement layer is transmitted, if the quality of the base layer is poor, the overall distortion will be high. In instances of an uncorrupted enhancement layer over a corrupted base layer, it will look as if noise has been added to the base layer due to the concealment efforts of the decoder in replicating parts of a frame from the previous frame to conceal the corrupted data.

Finally, the performance of the rate allocation algorithm is tested by simulating the transmission of the "Foreman" sequence over a wireless channel at 128 Kbps with the two-layer, high enhancement rate codec at 10 and 20 dB. The source and channel rates used to code the bitstream are given by the optimal algorithm. Table IV details the results. The results for the H.263 codec show that the distortion for the actual transmission is actually less than that given by the result of the optimization algorithm and for the 20 dB case is quite close to the uncorrupted results. Two representative frames from the transmission of the two-layer, high enhancement rates, "Foreman" sequence at 10 dB are shown in Figs. 11 and 12 with their uncorrupted counterpart. Overall, we conclude from the presented results that the rate allocation algorithm performs quite well in selecting the appropriate rates.

Finally, it should be noted that the structure of a video bitstream plays an important role in the error resiliency of source coded data. In H.263 efforts have been made to create an error resilient bitstream with the use of resync markers and specific Annexes. However, because of the extensive use of variable length codes (VLC) to code texture data, motion vectors, and mode information, the coded bitstream is especially susceptible to bit errors. In fact, in our simulations of video transmission, the majority of errors seen are due to VLC errors where corrupted bits cause illegal or incorrect VLC to be decoded. In the case of illegal VLCs being decoded, this leads to a complete

(a)

(b)

Fig. 12.   Decoded H.263 frame 51 of the "Foreman" sequence (a) without any channel errors and (b) 10 dB Rayleigh fading channel. The base layer is 42 Kbps protected with a channel rate of 1/2 and the enhancement layer is 28 Kbps with 2/3 rate channel protection.
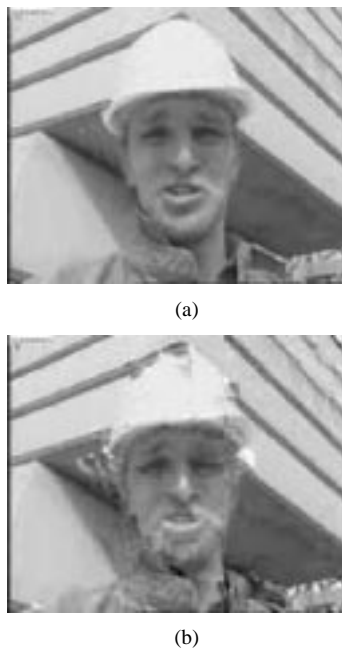
loss in synchronization at the decoder. The standard course of action for the decoder in this case is to reestablish synchronization at the next resync marker within the bitstream. In doing so all data between the place the illegal value occurred and the next resync marker is dropped since the decoder can no longer evaluate it. Within H.263 this resync either occurs at the beginning of a frame with the picture start code (PSC), or at the beginning of the next group of blocks (GOB). If Annex K (slice structure mode) is engaged, then the next resync can occur at the start of the next slice.

VI. CONCLUSIONS

In this work, we have developed an algorithm for the efficient transmission of scalable video over noisy channels. By formulating the problem of rate selection for scalable video as a dependent resource allocation problem, we proposed an optimal strategy for selecting the source and channel rates over a given number of layers given the characteristics of the source and channel. Given a target overall bitrate, the parameters of the transmission scheme, the channel, and the source codec, this algorithm provides appropriate breakdown of the rate into the subrates such that the overall end-to-end distortion is minimized. No other combination of rates totaling the given budget will result in lower distortion results. While the work presented has focused on point-to-point wireless communications, the formulation and solution is independent of the type of channel. All that is required is the channel model and the parameter(s) describing the state of the channel. In [29], the same joint source-channel coding algorithm is applied to a direct-sequence code-division-multiple-access (DS-CDMA) channel. Furthermore, while H.263 SNR scalability was used,

the proposed algorithm can be applied to any motion-compensated DCT-based scalable video coding technique such as MPEG-2 and other efficient scalable coding techniques [20]–[22]. Most of the previous work in the literature deals with rate allocation for subband-based scalable schemes. For a motion-compensated DCT-based scalable scheme, though, the layers are dependent. Thus, in this work, we define the enhancement layer distortion to be negative and also we condition the rate-distortion functions of a layer on the source coding rate selections for the previous layers.

In analyzing the results, we conclude that for reliable video transmission scalability is an effective tool. At low bitrates with a severely degraded channel, it can be seen that scalability provides a framework within which transmission can take place. With higher fidelity channels, it is also seen that scalable transmission generally outperforms nonscalable transmission. Beyond error resiliency, scalability clearly offers many more functionalities.

REFERENCES

[1] 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.   Englewood Cliffs, NJ: Prentice-Hall, 1998.

[2] C.-Y. Hsu, A. Ortega, and M. Khansari, "Rate control for robust video transmission over burst-error wireless channels," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 756–773, May 1999.

[3] G. Cheung and A. Zakhor, "Joint source/channel coding of scalable video over noisy channels," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, 1996, pp. 767–770.

[4] M. Bystrom and J. W. Modestino, "Combined source-channel coding schemes for video transmission over an additive white Gaussian noise channel," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 880–890, June 2000.

[5] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint source-channel coding for scalable video," in *Proc. Conf. Image and Video Communications and Processing*, 2000.

[6] R. E. Van Dyck and D. J. Miller, "Transport of wireless video using separate, concatenated, and joint source-channel coding," *Proc. IEEE*, vol. 87, pp. 1734–1750, Oct. 1999.

[7] H. Gharavi and S. M. Alamouti, "Multipriority video transmission for third-generation wireless communication systems," *Proc. IEEE*, vol. 87, pp. 1751–1763, Oct. 1999.

[8] M. Srinivasan and R. Chellappa, "Adaptive source-channel subband video coding for wireless channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1830–1839, Dec. 1998.

[9] M. Bystrom and J. W. Modestino, "Combined source-channel coding for transmission of video over a slow-fading Rician channel," in *Proc. Int. Conf. Image Processing*, 1998, pp. 147–151.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*.   New York: Wiley, 1991.

[11] J. Streit and L. Hanzo, "An adaptive discrete cosine transformed videophone communicator for mobile applications," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1995, pp. 2735–2738.

[12] L. Hanzo and J. Streit, "Adaptive low-rate wireless videophone schemes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, pp. 305–318, Aug. 1995.

[13] M. Khansari, A. Jalali, E. Dubois, and P. Mermelstien, "Low bit-rate video transmission over fading channels for wireless microcellular systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 1–11, Feb. 1995.

[14] K. Ramchandran, A. Ortega, K. M. Uz, and M. Vetterli, "Multiresolution broadcast for digital HDTV using joint source-channel coding," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 6–23, Jan. 1993.

[15] S. S. Prahdan and K. Ramchandran, "Efficient layered video delivery over multicarrier systems using optimized embedded modulations," in *Proc. IEEE Int. Conf. Image Processing*, 1997, pp. 452–455.

[16] I. Kozintsev and K. Ramchandran, "Multiresolution joint source-channel coding using embedded constellations for power-constrained time-varying channels," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1996, pp. 2345–2348.

[17] R. Aravind, M. R. Civanlar, and A. R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 426–435, Oct. 1996.

[18] Int. Telecommun. Union Recommend. H.263, "Video coding for low bitrate communications,", Jan. 1998.

[19] ISO/IEC 13 818-2 MPEG-2, "Information technology—Generic coding of moving pictures and associated audio information: Video,", 1995.

[20] L. P. Kondi and A. K. Katsaggelos, "An operational rate-distoriton optimal single-pass SNR scalable video coder," *IEEE Trans. Image Processing*, vol. 10, pp. 1613–1620, Nov. 2001.

[21] L. P. Kondi, "Low bit rate SNR scalable video coding and transmission," Ph.D. dissertation, Northwestern Univ., Evanston, IL, Dec. 1999.

[22] L. P. Kondi and A. K. Katsaggelos, "An optimal single pass SNR scalable video coder," in *Proc. IEEE Int. Conf. Image Processing*, 1999.

[23] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. Image Processing*, vol. 3, pp. 533–545, Sept. 1994.

[24] T. S. Rappaport, *Wireless Communications*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[25] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.

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

[27] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, pp. 389–400, Apr. 1988.

[28] F. Ishtiaq, "Scalable video compression and wireless video transmission at low bitrates," Ph.D. dissertation, Northwestern Univ., Evanston, IL, June 2000.

[29] L. P. Kondi, S. N. Batalama, D. A. Pados, and A. K. Katsaggelos, "Joint source-channel coding for scalable video over DS-CDMA multipath fading channels," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 2001, pp. 994–997.

Faisal Ishtiaq received the B.S., M.S., and Ph.D. degrees in electrical engineering in 1994, 1996, and 2000, respectively, all from Northwestern University, Evanston, IL.

He has worked in the area of low-bitrate scalable video coding and is currently a Research Engineer at Motorola, Inc., Schaumburg, IL. His areas of interest include algorithm development for low-bitrate video, scalability, error-resilient video coding, and video transmission over wireless and wireline channels. He has also been involved in the MPEG and H.26L video coding standardization efforts and has worked on numerous algorithm issues related to MPEG-4 and H.263/L video codecs.

Aggelos K. Katsaggelos (S'80–M'85–SM'92–F'98) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1979 and the M.S. and Ph.D. degrees, both in electrical engineering, from the Georgia Institute of Technology, Atlanta, in 1981 and 1985, respectively.

In 1985, he joined the Department of Electrical Engineering and Computer Science at Northwestern University, Evanston, IL, where he is currently Professor, holding the Ameritech Chair of Information Technology. He is also the Director of the Motorola Center for Communications. During the 1986–1987 academic year, he was an Assistant Professor with the Department of Electrical Engineering and Computer Science, Polytechnic University, Brooklyn, NY. His current research interests include image and video recovery, video compression, motion estimation, boundary encoding, computational vision, and multimedia signal processing and communications. He is a member of the Associate Staff, Department of Medicine, Evanston Hospital. He is an editorial board member of Academic Press, Marcel Dekker: Signal Processing Series, *Applied Signal Processing*, and the *Computer Journal*, and area editor for *Graphical Models and Image Processing* (1992–1995). He is the editor of *Digital Image Restoration* (Berlin, Germany: Springer-Verlag, 1991), co-author of *Rate-Distortion Based Video Compression* (Norwell, MA: Kluwer, 1997), and co-editor of *Recovery Techniques for Image and Video Compression and Transmission* (Norwell, MA: Kluwer, 1998). He is the coinventor of eight international patents.

Dr. Katsaggelos is an Ameritech Fellow and a member of SPIE. He is a member of the publication boards of the IEEE Signal Processing Society and the PROCEEDINGS OF THE IEEE, the IEEE TAB Magazine Committee, the IEEE Technical Committees on Visual Signal Processing and Communications, and Multimedia Signal Processing, and editor-in-chief of the *IEEE Signal Processing Magazine*. He has served as an associate editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1990–1992), a member of the steering committees of the IEEE TRANSACTIONS ON IMAGE PROCESSING (1992–1997) and the IEEE TRANSACTIONS ON MEDICAL IMAGING (1990–1999), a member of the IEEE Technical Committee on Image and Multidimensional Signal Processing (1992–1998), and a member of the Board of Governors of the IEEE Signal Processing Society (1999–2001). He has served as the General Chairman of the 1994 Visual Communications and Image Processing Conference, Chicago, IL, and as technical program co-chair of the 1998 IEEE International Conference on Image Processing, Chicago, and the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), and an IEEE Signal Processing Society Best Paper Award (2001).

Lisimachos P. Kondi (S'92–M'99) received the Diploma degree in electrical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1994 and the M.S. and Ph.D. degrees, both in electrical and computer engineering, from Northwestern University, Evanston, IL, in 1996 and 1999, respectively.

In August 2000, he joined the Department of Electrical Engineering, The State University of New York at Buffalo, as Assistant Professor. During the summer of 2001, he was a U.S. Navy-ASEE Summer Faculty Fellow at the Naval Research Laboratory, Washington, DC. His current research interests include video compression, wireless communications, joint source/channel coding, multimedia signal processing and communications, image restoration, resolution enhancement, and boundary encoding.