Joint Source Coding and Transmission Power Management for Energy Efficient Wireless Video Communications

Yiftach Eisenberg, *Student Member, IEEE*, Carlos E. Luna, *Student Member, IEEE*, Thrasyvoulos N. Pappas, *Senior Member, IEEE*, Randall Berry, *Member, IEEE*, and Aggelos K. Katsaggelos, *Fellow, IEEE*

***Abstract*—We consider a situation where a video sequence is to be compressed and transmitted over a wireless channel. Our goal is to limit the amount of distortion in the received video sequence, while minimizing transmission energy. To accomplish this goal, we consider error resilience and concealment techniques at the source coding level, and transmission power management at the physical layer. We jointly consider these approaches in a novel framework. In this setting, we formulate and solve an optimization problem that corresponds to minimizing the energy required to transmit video under distortion and delay constraints. Experimental results show that simultaneously adjusting the source coding and transmission power is more energy efficient than considering these factors separately.**

***Index Terms*—Error concealment, error resilience, expected distortion, optimal mode selection, power and rate control.**

I. INTRODUCTION

IN A WIRELESS setting, efficiently utilizing transmission energy is an important design consideration [1], [2]. Since most users of a wireless network are mobile, they must rely on a battery with a limited energy supply. Minimizing transmission energy can extend the lifetime of this battery. In covert communications, using the smallest amount of transmission energy to convey a message will decrease the likelihood of that message being intercepted. Reducing transmission energy can also decrease the interference between users sharing a wireless link, as well as increase the overall network capacity. Therefore, energy efficiency is a critical aspect of wireless communications.

In this paper, we introduce techniques for minimizing the energy needed to transmit a video sequence with an acceptable level of video quality and with tolerable delay. Two factors that directly impact this objective are the use of error concealment and resilience techniques at the source coding level, and the allocation of physical layer communication resources (such as the transmission power). Each of these factors has been well studied on its own. In this paper, we approach the problem of energy efficiency by jointly considering these factors in a common framework. We argue that the source coding and physical layer parameters should be adjusted simultaneously instead of adapting each independently. Experimental results show that this combined approach offers increased energy efficiency over approaches that focus on these factors separately.

Transmitting multimedia over unreliable networks, such as IP or cellular networks, has been an active field of research (see [3] for a recent survey). Work in this area has focused on various error resilience and error concealment techniques for minimizing the effects of losses [3]–[7]. These techniques attempt to encode the video sequence in ways that minimize the distortion at the receiver, given a statistical characterization of the channel errors. In [8], [9], the problem of optimal mode selection for transmission over lossy channels was considered. Knowledge of the decoder concealment strategy and the probability of packet loss were used by the encoder to select the coding mode for each macroblock that minimizes the expected distortion at the receiver. A similar approach was adopted in [10], [11]. The authors in [10] showed that the expected distortion at the receiver can be calculated recursively from frame to frame. In [11], the optimal placement of resynchronization markers is also considered. The techniques above assume that the probability of packet loss cannot be changed. In our work, by jointly considering the allocation of power at the physical layer, we incorporate the ability to control the loss characteristics of the channel. Therefore, we are able to adjust the reliability of the channel in response to variations in the source content.

At the physical layer, communication over wireless channels has also received considerable attention. Many of the physical layer techniques that have been considered for wireless channels can be classified as dynamic resource allocation techniques [12]. With these techniques, the transmitter can dynamically allocate communication resources, such as power and bandwidth, over time. The allocation may be based in part on any available knowledge of time-varying channel fading and interference. Transmitter power management has been an active field of research and was studied in [13]–[15], [1], [2]. Examples of transmitter power control are part of most emerging wireless standards [12]. One of the main assumptions made in developing these techniques is that the information bits are all equally important. Therefore, the emphasis is on performance measures such as throughput, assuming that all the bits must be delivered with an acceptably small probability of error. In video applications, this assumption does not hold because certain bits or packets can have greater significance. This significance depends on the source content as well as the particular

Manuscript received August 21, 2001; revised March 29, 2002.

The authors are with the Electrical and Computer Engineering Department, Northwestern University, Evanston, IL 60208 USA (e-mail: yeisenbe@ece.nwu.edu; carlos@ece.nwu.edu; pappas@ece.nwu.edu; rberry@ece.nwu.edu; aggk@ece.nwu.edu).

Publisher Item Identifier 10.1109/TCSVT.2002.800309.

source coding technique used. In our approach, the significance of each packet is used to simultaneously adjust the transmission power and the source coding parameters. Our goal is to effectively allocate power and bits in order to achieve the best possible video quality for the minimum amount of energy.

Recently there has been an increased interest in energy efficient wireless communications. Much of this work is fueled by the limited battery supply in mobile devices. In order to extend battery life, significant efforts have been made toward reducing the processing power required by the codec, decreasing transmission power, and increasing the efficiency of the power amplifier in mobile devices. In [16], saving transmitter power by judiciously suspending the communication device was studied. Here the strategy was to turn off the communication device completely whenever it was not needed, and only turn it on when required by the application.

Currently, there is an active field of research that focuses on minimizing transmission energy/power under quality of service requirements [17]–[20]. In order to efficiently utilize transmission energy, transmission rate adaptation along with transmission power adaptation has recently been studied [17], [18]. In [17], transmission rate adaptation is considered jointly with the source coding. This approach varies the transmission rate and source coding parameters per packet in order to minimize the total transmission energy needed to meet both distortion and delay constraints. An extension to this work, in which the number of macroblocks per packet is incorporated into the optimization, can be found in [21]. In this work, the authors assume that the transmission is nearly error free and therefore ignore distortion due to channel errors. The work presented here incorporates losses as well as the concealment strategy used at the decoder.

In [19], initial results where presented on minimizing the transmission energy required to meet expected distortion constraints. Knowledge of the concealment strategy and the relationship between transmission power and the probability of packet loss were used to simultaneously adjust the source coding parameters and transmission power per packet. Delay constraints imposed by the application were not considered in this work.

In [20], the authors considered the allocation of transmission rate between source and channel coding, as well as controlling the transmission power. Their goal was to use the minimal power to maintain a desired quality of service based on measurements such as throughput and error rate. They looked at both transmission power and the processing power required for source and channel coding. In this work, error concealment at the decoder was not considered. In addition, the authors did not consider varying the transmission power in order to provide a different quality of service to each packet. In our work, we incorporate the ability to change the protection given to each packet based on how difficult it is to conceal.

This paper is organized as follows. In the next section, we provide an overview of wireless video communication systems. In Section III, we present the problem formulation in detail. Section IV provides a general solution approach to the problem. We present an algorithm for optimally selecting the source coding parameters and transmission power per packet in Section V. Section VI presents experimental results. Conclusions are presented in Section VII.

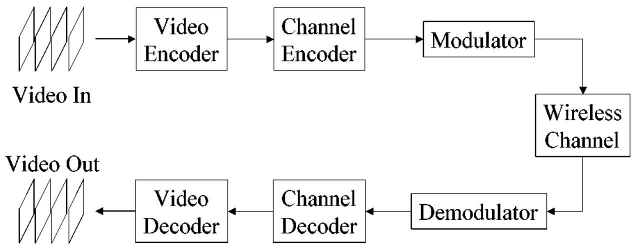

Fig. 1. Major components in a wireless video communication system.

II. Overview of Wireless Video Communications

In this section, we provide a brief high-level overview of a wireless video communication system. Fig. 1 highlights some of the major conceptual components found in such a system. The video encoder takes in an original video sequence and outputs an encoded version of that sequence. The two main objectives of the video encoder are to compress the video sequence and to make the encoded video resilient to errors. Compression reduces the number of bits used to describe the video sequence by exploiting both temporal and spatial redundancy contained in the sequence. The encoded video will be transmitted over a wireless channel that is lossy by nature. Therefore, the video sequence must be encoded in an error resilient way that minimizes the effects of losses on the decoded video quality.

The channel encoder adds redundancy to the bit stream via coding, in order to protect it from channel errors. Redundancy enables error detection and/or correction to be used at the channel decoder. The channel coding rate R_c is a measure of the redundancy added by the channel encoder and is defined as the number of video encoded bits per channel encoded bit.

The channel-encoded bit stream is then modulated and sent over the wireless channel. The modulation rate R_m is the number of channel-encoded bits per second transmitted across the channel. In addition to the modulation rate, the average transmission power used by the modulation scheme is an important quantity. The average transmission power directly affects the probability of error. The tradeoff for a lower probability of error is higher transmission power levels. In this paper, we show how this tradeoff can be managed in a way that minimizes the amount of transmission energy used to provide an acceptable level of video quality.

The wireless channel is modeled as a fading process that attenuates the transmitted signal plus an additive noise process. The fading process captures time variations in the channel response due to multipath interference, as well as shadowing and path loss. The noise process models thermal noise added in the receiver as well as other sources of interference. In a wireless setting, *channel state information* (CSI)—e.g., indicating the channel's fading level—may be available at the transmitter. Such information can be gained through direct feedback, detecting a pilot signal or measurements of the received signal in a duplex connection.

At the receiver, the demodulated bit stream is processed by the channel decoder, which performs error detection and/or

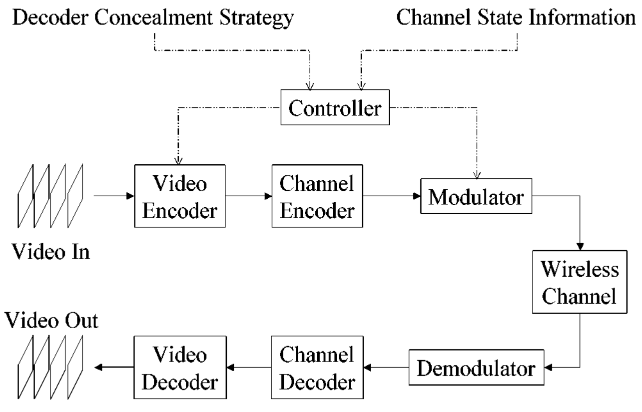

Fig. 2. System block diagram considered in this paper.

correction. Corrupt information can either be passed onto the video decoder or discarded. In this work, we assume that only error-free information is passed onto the video decoder and that corrupt data is considered lost. Instead of trying to determine where the error has occurred in a packet, we discard any packets in which an error is detected. We assume that the probability of an error being undetected is far smaller than the likelihood of a packet being lost due to a deep fade in the channel.

The video eecoder is responsible for reconstructing the received video sequence for display. Because some encoded information may have been lost, e.g., due to a deep fade in the channel, the video decoder must conceal any lost information. In this paper, we argue that knowledge of the error-concealment strategy used by the decoder is critical for the efficient allocation of both transmission power and source-coding resources.

III. Problem Formulation

We jointly consider adapting the source coding and physical layer parameters in order to efficiently utilize transmission energy while providing acceptable video quality. We consider the system shown in Fig. 2. In this setting the channel state information, as well as the decoder concealment strategy, are used to control the source-coding parameters and the transmission power. Our goal is to limit the amount of distortion in the received video sequence while using the minimal required transmission energy.

A. System Model

1) Source Coding: We consider a video application, where the video is encoded using a block-based motion-compensated video-coding technique (e.g., H.263 [22], MPEG-4 [23], etc.). With such a technique each frame is divided into M macroblocks. We assume that the macroblocks are numbered in scan line order and divided into groups called slices. Each slice is assumed to be independently decodable. This means that the video segment contained in each slice can be reconstructed independently of the other slices. After each slice is encoded, it is transmitted across a wireless channel as a separate packet. In the following, slice and packet will be used interchangeably. Let K be the number of packets in a given frame and k be the packet index.

For each macroblock, source-coding parameters such as the coding mode (inter/intra/skip) and the quantization step size are specified. We use μ^k to denote the source-coding parameters for all the macroblocks in the kth packet. Similarly, let B^k denote the total number of bits used to encode the kth packet. B^k is a function of the source coding parameters for that packet. We will use $B^k(\mu^k)$ to explicitly indicate this dependency. We assume that B^k also accounts for any additional overhead that is required for each packet.

2) Modulation and Channel Coding: In addition to the source-coding parameters, we assume that the average transmission power used for each packet can be adjusted. For the kth packet, let P^k be the average transmission power. We assume that the channel coding rate R_c and the modulation rate, R_m are fixed. Thus, each packet is transmitted at a fixed rate of $R = R_c R_m$ encoded video bits per second. The transmission delay for the kth packet is, therefore, B^k/R seconds.

B. Transmission Energy

Recall that we are interested in minimizing transmission energy. The total energy used to transmit all the packets in a frame is

$$E_{\mathrm{tot}} = \sum_{k=1}^{K} E^k = \sum_{k=1}^{K} \frac{B^k}{R} P^k \qquad (1)$$

where E^k is the energy used to transmit the kth packet. Notice that the total transmission energy is a function of the number of bits used to encode each packet and the power used to transmit them. This is one reason why we consider adapting the source coding and transmission power jointly.

C. Transmission Power and Probability of Packet Loss

The average transmission power used by a modulation scheme directly affects the probability of packet loss. By adjusting the transmission power we are able to control the level of protection we provide for each packet. We assume that a function f, relating the transmission power to the probability of packet loss, is known at the transmitter. This function can be determined from either empirical measurements or an analytical model of the wireless channel; we provide one example of this in Section VI. Let ρ^k denote the probability of loss for the kth packet. If average transmission power P^k is used for the kth packet, we have

$$\rho^k = f(P^k). \qquad (2)$$

Conversely, we can define the minimum transmission power required to achieve a desired probability of loss as

$$P^k = g(\rho^k). \qquad (3)$$

Assuming that f is strictly monotonic, then g will be the inverse of the function f.

The work in this paper is applicable to any mapping between P^k and ρ^k. What is important is that a function relating the transmission power to the probability of packet loss is known at the transmitter. Therefore, the work in this paper is not limited to a particular wireless communication scheme and can be applied to any system where the relationship between transmission power and probability of packet loss can be found. In the next section, we discuss how the quality of the decoded video is affected by the power allocated to each packet.

D. Expected Distortion

We consider the case where video quality is indicated by the expected distortion at the receiver, where the expectation is taken with respect to the probability of packet loss. The distortion between the original frame and the received frame depends on both the probability of packet loss and the source-coding parameters. The source encoder introduces distortion through quantization. The source encoder also affects the propagation of errors through mode selection [8]–[11]. As discussed above, adjusting the transmission power controls the probability of losing a packet.

We assume that the transmitter only knows the probability that a packet has arrived at the receiver. Thus, the distortion at the receiver is a random variable. Let $E[D^k]$ represents the expected distortion at the receiver for the kth packet. Given the probability of loss for the kth packet, ρ^k, the expected distortion for the kth packet can be written as

$$E[D^k] = (1-\rho^k)E[D_R^k] + (\rho^k)E[D_L^k] \tag{4}$$

where $E[D_R^k]$ is the expected distortion for the kth packet if the packet is received correctly at the decoder, and $E[D_L^k]$ is the expected distortion if it is lost. The reference frame at the encoder and at the decoder may be different because of packet losses. Therefore, due to temporal prediction, the distortion incurred if a packet is received (or lost) is also random. Hence, the expectation on the right-hand side of (4).

The expected distortion if a packet is received, $E[D_R^k]$, depends only on the source-coding parameters for that packet. In other words, if the source-coding parameters for the kth packet are fixed, then $E[D_R^k]$ is also fixed. We will use $E[D_R^k(\mu^k)]$ to explicitly indicate this dependency. The probability of loss for the kth packet ρ^k depends on the transmission power used for that packet, as in (2).

The expected distortion $E[D_L^k]$, if a packet is lost, depends on the concealment strategy used at the decoder. We assume that the encoder knows the concealment strategy, i.e., the encoder knows exactly how the decoder will conceal a macroblock if it is lost. Most concealment techniques today use information from neighboring macroblocks in order to conceal a lost macroblock [4], [5]. For example, many concealment strategies use temporal replacement based on the motion information of neighboring macroblocks [6], [7]. These techniques calculate a concealment motion vector for the lost macroblock based on the motion vectors of its neighboring macroblocks. The lost macroblock data is then replaced with the macroblock in the previous frame at the location defined by the concealment motion vector. In [24], it was shown that temporal replacement results in lower perceptual distortion than spatial interpolation. In Section VI, we present experimental results using a temporal concealment strategy.

The work presented here is applicable to any distortion metric. Thus the methods developed for minimizing transmission energy do not depend on how the distortion for a packet is measured. Therefore, as more sophisticated perceptual based distortion metrics are developed, such as the ones in [25] and [26], they may be used to calculate the distortion. Currently the mean squared error (mse) is commonly used to define the distortion. In [10] it is shown that the expected distortion for a frame, as in (4), can be calculated recursively if the mse distortion metric is used. This means that in order to calculate the expected distortion for each pixel in the current frame we only need to keep track of the first and second moment of each pixel value in the previous frame.

E. Optimization Formulation

Our goal is to control both the source-coding parameters and the transmission power in order to minimize the energy required to transmit a video frame at some acceptable level of quality and with tolerable delay. We can formally write this optimization as

$$\underset{\{\mu^k,\, P^k\}}{\text{minimize}} \quad E_{\text{tot}} = \sum_{k=1}^{K} \frac{B^k(\mu^k)}{R} P^k$$

subject to

$$E[D^k] = \begin{cases} D_0^k & \forall\, k\colon E[D_R^k(\mu^k)] \le D_0^k \le E[D_L^k] \\ E[D_L^k] & \forall\, k\colon D_0^k > E[D_L^k] \end{cases}$$

and

$$T_{\text{tot}} = \sum_{k=1}^{K} \frac{B^k(\mu^k)}{R} \le T_0 \tag{5}$$

where D_0^k is the acceptable expected distortion for the kth packet, T_{tot} is the total transmission delay for the frame, and T_0 is the maximum amount of time that can be used to transmit the entire frame. Recall that $B^k(\mu^k)P^k/R$ is the transmission energy for the kth packet. Thus, the objective in (5) is to minimize the energy used to transmit the entire frame.

The acceptable level of quality for each packet may vary based on the application. For example, in a surveillance scenario, video packets containing the object being tracked may require more stringent distortion constraints than the background. Therefore our approach allows different packets to have different levels of acceptable distortion, D_0^k. The acceptable distortion for each packet must be specified, and therefore, the D_0^k's are known constants in our formulation.

When D_0^k is between $E[D_R^k(\mu^k)]$ and $E[D_L^k]$, we constrain the expected distortion for that packet to be equal to D_0^k. We assume that there exists a coding option for each packet that has $E[D_R^k(\mu^k)] \le D_0^k$. If this condition is not satisfied, then the problem becomes infeasible. Also, $E[D_R^k(\mu^k)]$ must be less than $E[D_L^k]$. This makes sense because it says that the distortion resulting from a packet being lost and concealed ($E[D_L^k]$) must be greater than the distortion if the packet is received ($E[D_R^k(\mu^k)]$). If this were not true, then it would be better not to send the packet and instead let the decoder conceal it.

When D_0^k is greater than $E[D_L^k]$, it means that the expected distortion if the packet is lost is below the acceptable distortion level. Therefore, we need not transmit this packet and thus the expected distortion for this packet equals $E[D_L^k]$. This is a very important special case, similar to the "skip" mode in the MPEG and H. standards. We will therefore refer to this as the *generalized skip mode.* The generalized skip mode is an option that allows the transmitter to save both time and energy by not transmitting a packet. In the previously mentioned standards, a macroblock that is encoded using the skip mode is reconstructed at the decoder by copying the spatially equivalent macro block in the previous frame to the current macroblock location. By using the generalized skip mode, the encoder forces the decoder to conceal all the macroblocks in a packet using information from neighboring packets. For example, if there is a high correlation between the motion in a group of neighboring macroblocks, then the encoder could chose not to transmit selected packets if their expected distortion when concealed is acceptable. This allows the transmitter to exploit the correlation between the neighboring macroblocks in order to allocate more time and energy to other packets.

In many applications, such as video conferencing and streaming, there is a limited amount of time by which the video sequence must arrive at the decoder [27], [28]. We assume that processing and propagation delays are constant and can be ignored in this formulation. We only need to concern ourselves with the transmission delay. Typically, a higher level rate controller will assign a bit budget to each frame in order to meet any delay constraints imposed by the application. The video encoder must then encode each frame so that it meets this bit budget constraint. In our work, we assume that a similar delay controller assigns a transmission delay constraint to each frame. This means that each frame must be transmitted within T_0 seconds. Note that the value of T_0 can vary from frame to frame. Since the transmission rate R is fixed, the transmission delay constraint translates directly to a bit budget constraint for the frame. Therefore, $T_0 R$ is the maximum number of bits that can be used to encode a given frame. If a frame uses less than its maximum transmission time, the excess time may be allocated to future frames. In future work, we plan to incorporate rate adaptation into the optimization. This means that R_c, R_m, and therefore R could vary from packet to packet. Thus, a bit budget constraint is no longer applicable. This is why we use a delay constraint instead of a bit budget constraint per frame. Initial work on using transmission rate adaptation for energy efficient wireless video communications can be found in [17].

We assume that a higher level controller assigns an expected distortion constraint to each packet D_0^k, and a transmission delay constraint T_0 to each frame. Our objective is then to select the source-coding parameters μ^k and the transmission power per packet P^k that minimize the amount of transmission energy needed to meet the quality and delay constraints. In the next section, we present methods for meeting this objective.

IV. Minimizing Transmission Energy

In this section, we present a solution to the minimum transmission energy optimization problem in (5). We present methods for selecting the optimal source-coding parameters and transmission power per packet that minimize the total transmission energy for a video frame. First we use Lagrangian relaxation for the delay constraint. Then, we discuss how the distortion constraint couples the transmission power to the source-coding parameters and thus allows us to reformulate our problem as an optimal source-coding problem.

A. Lagrangian Relaxation

In order to meet the transmission delay constraint we introduce a Lagrange multiplier λ and solve the following relaxed problem:

$$\underset{\{\mu^k, P^k\}}{\text{minimize}} \quad J_{\text{tot}} = \sum_{k=1}^{K} \frac{B^k(\mu^k)}{R} P^k + \lambda \sum_{k=1}^{K} \frac{B^k(\mu^k)}{R}$$

subject to

$$E[D^k] = \begin{cases} D_0^k, & \forall\, k\colon E[D_R^k(\mu^k)] \le D_0^k \le E[D_L^k] \\ E[D_L^k], & \forall\, k\colon D_0^k > E[D_L^k], \end{cases} \tag{6}$$

where J_{tot} is the cost function to be minimized, and the distortion constraint is identical to the one in (5). The cost in (6) is comprised of the transmission energy for the frame, plus the total transmission time multiplied by λ. By appropriately choosing λ, the solution to (5) can be obtained within a convex-hull approximation by solving (6) [29]. In this way, we are able to solve a simpler relaxed problem a few times instead of solving a hard problem once.

Notice that λ acts like a constant power multiplying the transmission time for the frame. Thus, the second term in the cost function can be interpreted as a constant energy that depends only on the transmission time for the frame. Therefore, as λ increases, coding options with lower transmission times become more favorable than ones that take longer to transmit.

B. Cost Function Redefined Using the Distortion Constraint

Recall that the expected distortion for the kth packet is defined in (4). Therefore, by setting the expected distortion for the kth packet equal to the expected distortion constraint in (6) and solving for ρ^k, the probability of loss for the kth packet can be expressed as

$$\rho^k = \begin{cases} \dfrac{D_0^k - E[D_R^k(\mu^k)]}{E[D_L^k] - E[D_R^k(\mu^k)]}, & \text{if } E[D_R^k(\mu^k)] \le D_0^k \le E[D_L^k] \\ 1, & \text{if } D_0^k > E[D_L^k]. \end{cases} \tag{7}$$

Equation (7) is used to calculate the exact probability of loss required for each packet to meet its expected distortion constraint. Recall that if D_0^k is greater than $E[D_L^k]$, we do not transmit the kth packet, and therefore set ρ^k equal to one.

We can substitute (7) into (3) in order to express the transmission power required for the kth packet to meet its distortion constraint as

$$P^k = \begin{cases} g\left(\dfrac{D_0^k - E[D_R^k(\mu^k)]}{E[D_L^k] - E[D_R^k(\mu^k)]}\right), & \text{if } E[D_R^k(\mu^k)] \le D_0^k \le E[D_L^k] \\ 0, & \text{if } D_0^k > E[D_L^k]. \end{cases} \tag{8}$$

Therefore, the required power for the kth packet depends on the source-coding parameters for that packet, which determine $E[D_R^k(\mu^k)]$, and the distortion incurred if the packet is lost, $E[D_L^k]$.

By combining (1) and (8), the required transmission energy for the kth packet is given by

$$E^k = \begin{cases} \dfrac{B^k}{R}\, g\left(\dfrac{D_0^k - E[D_R^k]}{E[D_L^k] - E[D_R^k]}\right), & \text{if } E[D_R^k] \le D_0^k \le E[D_L^k] \\ 0, & \text{if } D_0^k > E[D_L^k]. \end{cases} \tag{9}$$

Equation (9) specifies exactly how much energy is needed to transmit each packet in order to meet its expected distortion constraint.

We can now redefine the cost for transmitting all the packets in the frame to be

$$J_{\text{tot}} = \sum_{k=1}^{K} J^k \tag{10}$$

where J^k is the cost for transmitting the kth packet given by

$$J^k = \begin{cases} \dfrac{B^k}{R}\left[g\left(\dfrac{D_o - E[D_R^k]}{E[D_L^k] - E[D_R^k]}\right) + \lambda\right], & \text{if } E[D_R^k] \le D_o \le E[D_L^k] \\ 0, & \text{if } D_o > E[D_L^k]. \end{cases} \tag{11}$$

Recall that the function $g(\bullet)$ is known at the transmitter and is used to relate the transmission power to the probability of loss. Since the transmission rate is fixed, R is a constant. The expected distortion constraint for each packet, D_0^k, and λ are specified, and are therefore constants in the equations above. Thus, the only variables are $B^k(\mu^k)$, $E[D_R^k(\mu^k)]$, and $E[D_L^k]$. Note that due to the concealment strategy, $E[D_L^k]$ may be a function of the source-coding parameters and transmission power used for neighboring packets in the frame.

C. Coupling Power to Source-Coding Parameters

We now show how the distortion constraint couples the transmission power per packet to the source-coding parameters for the frame. This enables us to reduce our overall problem of finding the optimal source coding and transmission power allocation for the frame into a problem of finding the optimal source-coding parameters that minimize the total cost for the frame.

We assume that the concealment strategy used by the decoder is *spatially causal*. This means that the decoder will only use information from previously received packets in order to conceal a lost packet. Therefore, $E[D_L^k]$ can only depend on how previously transmitted packets were encoded and the likelihood that they arrived correctly at the decoder. Future packets will not affect the concealment of the current packet.

For the first packet in a frame, there are no previous packets in that frame that can be used to help conceal it. Therefore, the expected distortion if the first packet is lost, $E[D_L^1]$, must be a known constant. Given $E[D_L^1]$, (11) can be used to calculate the exact cost for transmitting the first packet based only on its source-coding parameters, i.e., $J^1(\mu^1)$. When the source-coding parameters for the first packet are fixed, the transmission power for that packet is also fixed using (8). In other words, for a given μ^1, (8) tells us exactly how much transmission power to use for the first packet in order to meet its expected distortion constraint.

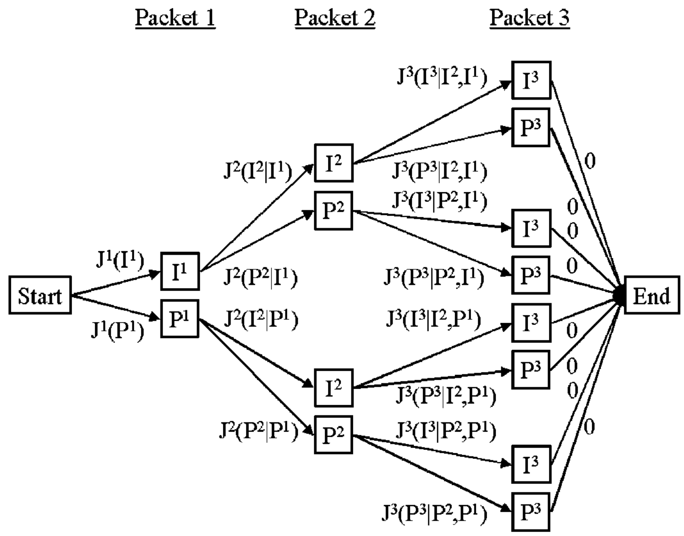

Fig. 3. Source-coding tree for a frame containing three packets, where each packet can be encoded in one of two ways, I or P. I^k represents the kth packet coded using the first option and P^k represents the kth packet coded using the second option. The weight of each branch, $J^k(\bullet)$, represents the cost for coding the kth packet given the coding sequence for the previous packets.

Once the source coding—and therefore the transmission power—for the first packet are fixed, $E[D_L^2]$ must become a known constant since it can only depend on how the first packet is encoded and transmitted. When $E[D_L^2]$ is known, (11) can again be used to determine the cost for the second packet based only on its source-coding parameters μ^2. The third packet in the frame may depend on the previous two packets for concealment. If the source-coding parameters for the first and second packet are fixed, $E[D_L^3]$ must be a known constant, and therefore the cost for the third packet is uniquely defined by its source-coding parameters. We can continue to recursively calculate the cost to transmit all the packets in the frame based solely on the source-coding parameters for the previous packets.

A source-coding sequence specifies the source-coding parameters for a group of packets. As shown above, the source-coding sequence used for the frame uniquely defines the transmission power and energy as well as the total cost for the frame. Therefore, in order to minimize the total cost, J_{tot}, we only need to find the source-coding sequence that has the smallest cost. In summary, we have shown that the distortion constraint in (6), along with a spatially causal concealment strategy, allows us to simplify our problem of finding both the optimal source-coding parameters and transmission power per packet into a problem of finding only the optimal source-coding sequence for the frame.

D. Tree Structured Optimization

All the possible source-coding sequences can be thought of as a *source coding tree,* as shown in Fig. 3. Each node in the tree represents a particular source coding choice for a given packet. The weight of each branch represents the cost of coding a packet given the source-coding sequence for the previous packets. The terminal cost for every sequence is zero. We must find the least costly path from the start to the end. In the example shown in Fig. 3, there are three packets in the frame, and each packet can be encoded in one of two ways, e.g., in one of two modes. We denote the kth packet encoded using the first option as I^k, and using the second option as P^k.

Since all the previous packets may affect the cost for the current packet, this is a tree-structured dependent optimization problem [30]. Shortest-path algorithms can be used to find the optimal source-coding sequence. The drawback of such an approach is that the entire source-coding tree, which grows exponentially with the number of packets, must be explicitly constructed before these algorithm can be applied. Next, we present an alternative algorithm that finds the optimal source-coding sequence without having to explicitly construct the entire tree, thereby reducing the computational complexity in finding the optimal answer.

V. Optimal Source-Coding Algorithm

In this section, we introduce an algorithm that finds the optimal solution to the minimum transmission energy formulation (5). We conclude with a simple example to demonstrate how our algorithm works. The concealment strategy used by the decoder introduces dependencies between packets through the calculation of $E[D_L^k]$. This means that the cost J^k for a given packet depends on how its neighboring packets are encoded and transmitted. Our algorithm exploits the possible limitations of the dependencies between packets in order to arrive at the optimal answer without checking every possible source-coding sequence.

The source-coding tree for the frame is constructed in reverse. If a group of paths originating at a common node do not depend on what coding choices are made for previous packets, their path weights must be known constants. Therefore, pruning between paths is only done when the cost of the paths originating at a common node are constant. When all the packets have been incorporated into the tree, the cost for each of the remaining coding sequences must be a known constant, as discussed in Section IV-C. Therefore, when all the packets are incorporated into the tree, we are able to prune down to one coding sequence. This coding sequence is the one that minimizes the total cost for the frame. Remember that the transmission power used for each packet in this coding sequence is defined by (8). Therefore, the algorithm will always converge to the optimal source-coding sequence and power per packet that minimize the total cost, J_{tot}. If the resulting transmission time is greater than T_0, then λ is adjusted until the delay constraint is met. Like all Lagrangian relaxations, finding the value of λ that produces the optimal answer to the unrelaxed problem (5) can be performed in a variety of ways [29]. Once a frame has been encoded and transmitted, the next frame can be optimized.

Next, we formally present our algorithm. This is followed by an example to help illustrate how the algorithm works.

Given: T_0, $D_0^k \ \forall\, k$, λ, $g(\bullet)$, and the concealment strategy.

1) Define $E[D_L^k]$ for the concealment strategy used at the decoder.
2) Determine the dependencies between the packets in a frame based on the concealment strategy.
3) Initialize: Let $k = K$. (start with the last packet)
4) Construct the sub-graph of the source-coding tree which contains the following nodes: i) all of the source-coding choices for the kth packet, ii) all of the packets that the kth packet depends on for concealment, and iii) the surviving source-coding combinations for the packets already considered.
5) For any group of paths in the sub-graph originating at a common node (root) and whose costs are only functions of packets k through K, keep as a survivor the path with the smallest total cost.
6) If $k \neq 1$, decrement k by one and go to step 4. Otherwise stop.

A. Example Implementation of the Algorithm

In this example, we demonstrate how our algorithm is used to encode and transmit a video frame with the minimum required transmission energy. We assume that each transmitted packet contains only one macroblock. Therefore, each macroblock is independently encoded by defining a slice to be a single macroblock. This packetization scheme has a low coding efficiency but helps illustrate the concepts introduced in this paper. Assume that for each macroblock we can select the coding mode, Intra (I) or Inter (P). Therefore, the coding mode is the source-coding parameter that must be specified for each packet.

For the concealment strategy, we assume that if a macroblock is lost, the motion vector of the spatially neighboring macroblock to the left is used as the concealment motion vector for the lost macroblock. If the previous macroblock is also lost, then the concealment motion vector for the current macroblock is set to zero. If a macroblock is on the left edge of the frame, then the concealment motion vector for that macroblock is also set to zero.

Step 1: For this concealment strategy we define $E[D_L^k]$ to be

$$E[D_L^k] = \begin{cases} (1-\rho^{k-1})E[D_C^k] + (\rho^{k-1})E[D_Z^k], & \text{if not on left edge} \\ E[D_Z^k], & \text{if on left edge} \end{cases} \tag{12}$$

where ρ^{k-1} is the probability that the neighboring macroblock to the left is lost, $E[D_C^k]$ is the expected distortion if the concealment motion vector equals the motion vector of the previous macroblock, and $E[D_Z^k]$ is the expected distortion if the concealment motion vector equals zero. For all the macroblocks on the left edge of the frame, $E[D_L^k] = E[D_Z^k] =$ a constant.

Step 2: Based on the concealment strategy described above, we draw the dependencies between packets in the frame, as shown in Fig. 4. This illustrates which neighboring packets are used to help conceal each packet if it is lost. For this concealment strategy the dependencies between packets do not cross between rows of packets. Therefore, each row of macroblocks can be independently optimized. Note that Fig. 3 is the source-coding tree for one row of packets in this example.

Steps 3 and 4: We demonstrate how to apply our algorithm to the first row of packets. The optimization process for the other rows in the frame is identical. We construct the sub-graph containing all the ways to encode the last packet in the first row

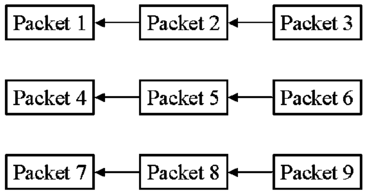

Fig. 4. Dependencies between packets due to the concealment strategy. This figure shows which packets are used to conceal each packet if it is lost. For example, if packet 3 is lost, then the motion vector of packet 2 is used to help conceal packet 3.

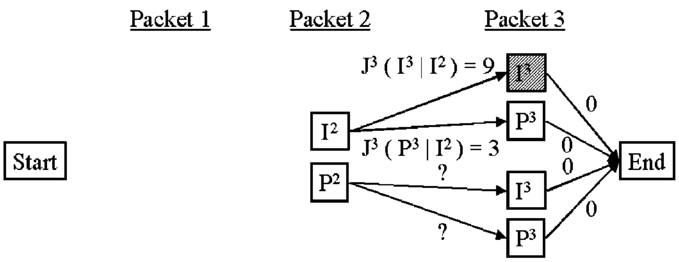

Fig. 5. Sub-graph containing all possible source-coding combinations for the last packet and the packet it depends on for concealment. A source-coding option that has been pruned is indicated with dashed lines through it. Branch weights that are unknown are identified with a question mark.

(packet 3) and the packet it depends on for concealment (packet 2), as shown in Fig. 5.

Step 5: Recall that $E[D_L^k]$ depends on the motion vector used for the previous packet and its probability of loss ρ^{k-1}. Therefore, if the second packet is coded as Inter (P) with nonzero motion vector, then $E[D_L^3]$ is not known until ρ^2 is known. In this case, we are unable to determine the costs $J^3(\mathrm{I}^3|\mathrm{P}^2)$ or $J^3(\mathrm{P}^3|\mathrm{P}^2)$ until we know the probability of loss for the second packet, ρ^2.

On the other hand, if the second macroblock is coded as Intra (I), then the concealment motion vector for the third macroblock equals zero regardless of the probability of loss for the second packet, and $E[D_L^3] = E[D_Z^3] =$ a constant. Thus, if the previous packet is coded as Intra, then the cost for coding the current packet depends only on how it is coded. As shown in Fig. 5, when the second packet is coded as Intra, I^2, we keep as a survivor only the coding option P^3 because it has a lower cost than I^3.

Steps 5 and 6: We now construct the sub-graph containing the surviving source-coding combinations for packets 2 and 3, and all the ways to code packet 1, as shown in Fig. 6. We are able to prune between all the paths leaving I^1, since the costs of all the paths leaving this node can be calculated. In this example, the least costly coding sequence for packets two and three is $(\mathrm{P}^2,\ \mathrm{P}^3)$ when the first packet is coded as Intra. Notice that this would also be the least costly way to encode the last two packets if the first packet was coded as Inter (P) with a motion vector equal to zero.

Step 6: Once all three packets have been incorporated into the graph, the cost of all the surviving source-coding sequences can be calculated. We can therefore prune between all the remaining paths to find the optimal way to encode each

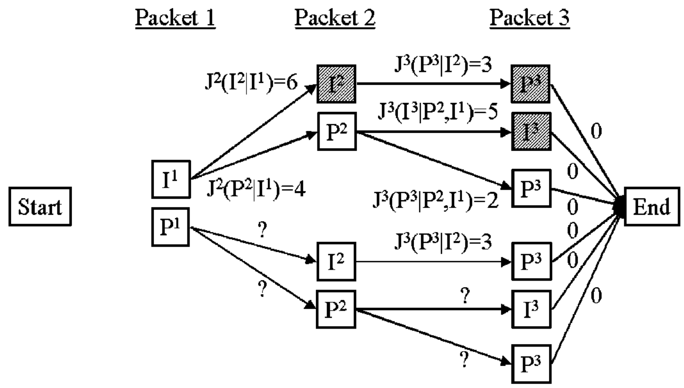

Fig. 6. Sub-graph containing the surviving source-coding combinations for packets 2 and 3, and all the ways to encode packet 1.

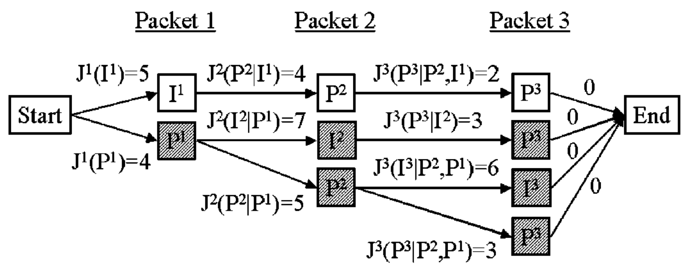

Fig. 7. Final sub-graph. The optimal source-coding sequence is $(\mathrm{I}^1,\ \mathrm{P}^2,\ \mathrm{P}^3)$ and its total cost is 11.

packet so that the total cost is minimized. In this example the least costly source-coding sequence is $(\mathrm{I}^1,\ \mathrm{P}^2,\ \mathrm{P}^3)$, as shown in Fig. 7. Notice that if we had started with the first packet and selected the minimum cost way to encode it without considering its effects on the cost of future packets, we would have selected $(\mathrm{P}^1,\ \mathrm{P}^2,\ \mathrm{P}^3)$, which is sub-optimal. Our approach of constructing the source-coding tree in reverse and pruning between paths with constant costs guarantees that we will choose the optimal answer.

The number of sequences that must be considered depends on the concealment strategy, the constraints chosen for the frame, and on the video sequence itself. For the concealment strategy in this example, each row of macroblocks can be independently optimized. We also noticed that we are able to prune between all the paths leaving a macroblock that is coded as Intra, coded as Inter with zero motion vector, or is not transmitted. All of the scenarios described above will make $E[D_L] = E[D_Z] =$ a constant for the next macroblock. Therefore, if there is little motion between frames, the optimization is faster because most of the macroblocks will have zero motion vectors when coded as Inter.

VI. Experimental Results

In this section, we present experimental results that demonstrate the advantages of simultaneously adjusting the source coding and transmission power in wireless video communications. As an alternative, we consider an approach that optimizes the source coding and the transmission power independently. We demonstrate that, as expected, jointly optimizing the source coding and transmission power is more energy efficient than optimizing each independently. We show that our approach requires less energy to achieve the same level of quality, and vice versa, provides higher quality video for the same amount of transmission energy.

A. Fixed Packet Loss Approach

As an alternative to the minimum energy (ME) approach presented in (5), we consider a scheme that maintains a fixed probability of packet loss and optimally allocates bits in order to minimize the maximum expected distortion. We will refer to this scheme as the fixed packet loss (FPL) approach. In this approach, the source-coding parameters and the transmission power are independently optimized. Because the source encoder and the transmitter operate independently, the relative importance of each packet, i.e., their contribution to the total distortion, is unknown to the transmitter. Therefore, the transmitter treats each packet equally. In order to efficiently utilize transmission energy, the transmitter uses the minimum amount of transmission power in order to provide a guaranteed quality of service to the source encoder via a constant probability of packet loss.

The source encoder is unable to change the probability of loss for each packet because it cannot control the transmission power. Therefore, the goal of the source encoder is to provide the best video quality for a given probability of packet loss. In these experiments, we use the maximum expected distortion in a frame as the measure of video quality. Thus, we formulate this problem as a minimum–maximum distortion approach. The problem can be written as

$$\underset{\{\mu^k\}}{\text{minimize}} \quad D_o = \max_{k\in[1,\ldots,K]} \{E[D^k]\}$$

subject to

$$T_{\text{tot}} = \sum_{k=1}^{K} \frac{B^k(\mu^k)}{R} \leq T_0,$$

where

$$E[D^k] = \begin{cases} (1-\rho)E[D_R^k] + (\rho)E[D_L^k] \\ \qquad\qquad \forall\, k\colon E[D_R^k] \leq D_o \leq E[D_L^k] \\ E[D_L^k] \qquad \forall\, k\colon D_o > E[D_L^k] \end{cases}$$

and

$$E_{\text{tot}} = \sum_{k=1}^{K} \frac{B^k(\mu^k)}{R}\, P \tag{13}$$

where D_o is the maximum expected distortion in the frame and P is the minimum transmission power required to provide a constant probability of packet loss ρ. The formulation in (13) is similar to the one in (5). Both formulations have the same transmission delay constraint per frame. Notice that the *generalized skip mode* enables both formulations to not allocate any resources to packets whose expected distortion when lost, $E[D_L^k]$, is below the maximum expected distortion D_o. The important distinction between the two formulations is that in (5), the probability of loss can be adjusted per packet by controlling the transmission power, and in (13), the probability of packet loss is fixed.

B. Experimental Set-Up

1) Source Coding: We consider the packetization scheme and concealment strategy described in Section V-A. Therefore, each macroblock is independently encoded and transmitted across the channel as a separate packet. If a macroblock is lost, the decoder will conceal it using the motion vector of the spatially previous macroblock to the left. If the macroblock to the left is also lost, then the decoder will use zero motion vector concealment for the current macroblock (see Section V-A). We consider the "Foreman" sequence in QCIF format encoded at 15 frames per second using an MPEG-4 codec. We consider a limited number of quantizers for each macroblocks. The available Intra mode quantizers are of step size 6, 12, 18, 24, and the available Inter mode quantizers are of step-size 6, 12, skip, and the *generalized skip mode*. We assume that the first frame in the sequence is coded as Intra with a quantization step size of 15 and that enough transmission power is used so that it arrives correctly at the decoder. This is done so that the initial conditions of all the experiments are identical. If we were to consider losses in this Intra frame, a spatial concealment strategy would be used.

In these experiments, video quality is measured by the maximum expected distortion in a frame. We assume that each macroblock is equally important. Therefore, in the ME approach, each transmitted packet is constrained to have the same maximum expected distortion, i.e., we set $D_o^k = D_o$ for all k in (5). We further assume that each video frame is equally important and thus constrain the maximum expected distortion to be the same for each frame in the video sequence. For the FPL approach, the objective is to minimize the maximum expected distortion per frame given a fixed probability of packet loss.

We consider a "real time" application where the maximum allowable transmission delay is one frame duration. Since the video sequence is encoded at 15 frames per second, the maximum allowable transmission delay per frame is $T_0 = 67$ ms.

2) Channel Model: We consider the case where each packet is sent over a narrow-band slowly-fading channel with additive white Gaussian noise. In this case the received signal-to-noise ratio (SNR) for the kth packet is given by $H^kP^k/(N_0W)$, where H^k is a random variable representing the channel's fading, P^k is the transmission power for the kth packet, and N_0W represents the noise power. For this example, we assume that the channel fading H^k stays fixed during the transmission of an entire packet, but can vary between packets. We assume that $\{H^k, k = 1, \ldots, K\}$ is an independent and identically distributed sequence of random variables. We assume that the distribution of these random variables is known at the transmitter, but the actual realization is not known. For example, this knowledge can be gained from measurements of a pilot signal broadcast by the receiver or from direct feedback [12].

After transmission, we assume that each packet either arrives error free or is dropped due to a channel fade. We model the probability of packet loss in the capacity versus outage framework introduced in [31]. That is, we assume that a packet is received error-free if the transmission power is large enough so that the channel capacity for a given fading realization is greater than R bits per second, i.e., $\rho^k = \Pr(C(H^kP^k) \leq R)$, where $C(x)$ is the Shannon capacity of a bandlimited AWGN channel with received power H^kP^k. We consider a Rayleigh fading channel, so that H^k will have an exponential distribution and mean $E[H]$. Recall that the expected channel state $E[H]$ is fixed for each packet and is known at the transmitter. In this case, the relationship between ρ and P^k can be expressed as

$$\rho^k = 1 - e^{-G/P^k}$$

where

$$G = \frac{N_0W}{E[H]}\left(2^{R/W} - 1\right). \tag{14}$$

The exact power needed to achieve a desired probability of loss can be calculated by rewriting (14) as

$$P^k = \frac{-G}{\ln(1-\rho^k)}. \tag{15}$$

Therefore, for this channel model $g(\rho^k) = -G/\ln(1-\rho^k)$. Recall that in the ME approach, we use (7) to calculate the exact probability of loss required for the kth packet to meet its expected distortion constraint. In other words, (7) is used to calculate the required probability of packet loss, and (15) is used to calculate how much transmission power is needed to achieve this desired probability of loss.

In our experiments, $N_0W/E[H] = 6$ W, $W = 5$ MHz, and $R = 225$ kbits/s. These values are similar to the ones being proposed for next generation wireless standards [12]. Since $R = 225$ kbits/s, the bit budget per frame is $T_0R = 15$ kbits.

C. Results

We compare the ME approach to the FPL approach. Recall that in the ME approach, the maximum expected distortion per frame (D_o) is given, and the objective is to minimize the transmission energy. In the FPL approach, the objective is to minimize the maximum expected distortion per frame for a given probability of packet loss.

In Fig. 8, the average transmission energy versus the average expected distortion is shown for the "Foreman" sequence. The average transmission energy for the sequence is the average transmission energy per frame. The average expected distortion for the video sequence if defined as the average maximum expected distortion per frame, i.e., the average D_o per frame. By varying the distortion constraint, we obtain the curve for the ME approach. By varying the probability of packet loss, we obtain the results for the FPL approach. We see that for the same distortion, the ME approach uses less transmission energy than the FPL approach. For example, in order to achieve an average expected distortion of 132 mse, the ME approach requires an average transmission energy of 0.0738 Joules while the FPL approach requires 0.1719 Joules. In this case, the ME approach uses 57% less energy than the FPL approach. Similarly, we notice that for the same amount of transmission energy, the ME approach provides higher quality video. Using an average of 0.05 Joules of energy per frame, the FPL approach has an average expected distortion that is nearly twice as large as the ME approach.

Fig. 8 illustrates the relationship between transmission energy and the expected video quality at the decoder for the entire video sequence. In Fig. 9(a) and (b), we show how the maximum expected distortion and the transmission energy vary from frame to frame. We compare the ME approach with $D_o = 132$ (mse) and the FPL approach with $\rho = 0.0494$. In this example, both approaches have the same average expected distortion (132 mse), but the ME approach has an average transmission energy that is 57% smaller. As shown in Fig. 9(a), the ME approach maintains a constant level of video quality across all the frames. On the other hand, the maximum expected distortion for the FPL approach varies greatly throughout the sequence. The peaks in the distortion correspond to periods of high activity in the video sequence. For example, from about frame 80 to 110, there is a scene change where the camera pans from the foreman to the construction site. During this period, the maximum expected distortion increases significantly for the FPL approach. During high activity periods, it may be more difficult to conceal lost packets because the correlation between consecutive frames is smaller. This suggests that more protection should be given to frames with high activity. Our ME approach does exactly that.

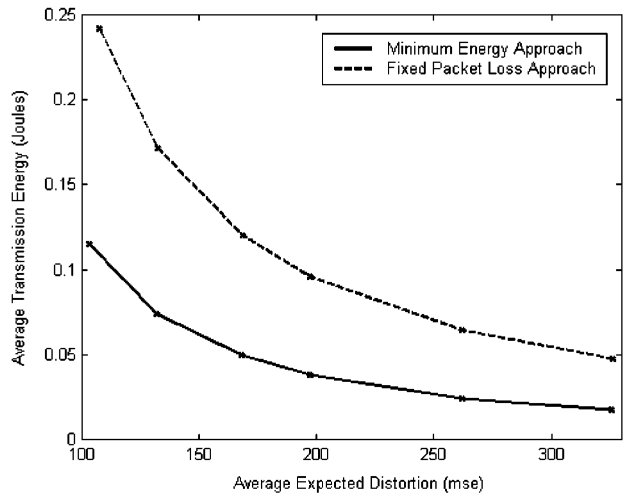

Fig. 8. Average transmission energy versus average expected distortion per frame for the "Foreman" sequence.

In the ME approach, transmission power can be adjusted in order to control the probability of packet loss. Therefore, in periods of high activity the ME approach can increase the transmission power, and thus the transmission energy, in order to increase the likelihood that these frames will arrive at the decoder correctly, as shown in Fig. 9(b). When there is little activity in the sequence, the ME approach can use less transmission energy in order to maintain the same expected video quality. This enables the ME approach to save energy for when there are larger changes in the video sequence.

In the FPL approach, the probability of packet loss is fixed. Therefore, the FPL approach is unable to reduce the transmission power during periods of low activity and increase it during high activity periods. This is due to the fact that the video encoder and the transmitter act independently in the FPL approach. Therefore, the ME approach is better able to adapt to changes in the source content by varying the transmission energy per frame in relation to the activity level, as shown in Fig. 9(b).

In addition to allocating energy differently between frames, the ME approach allocates resources differently between the

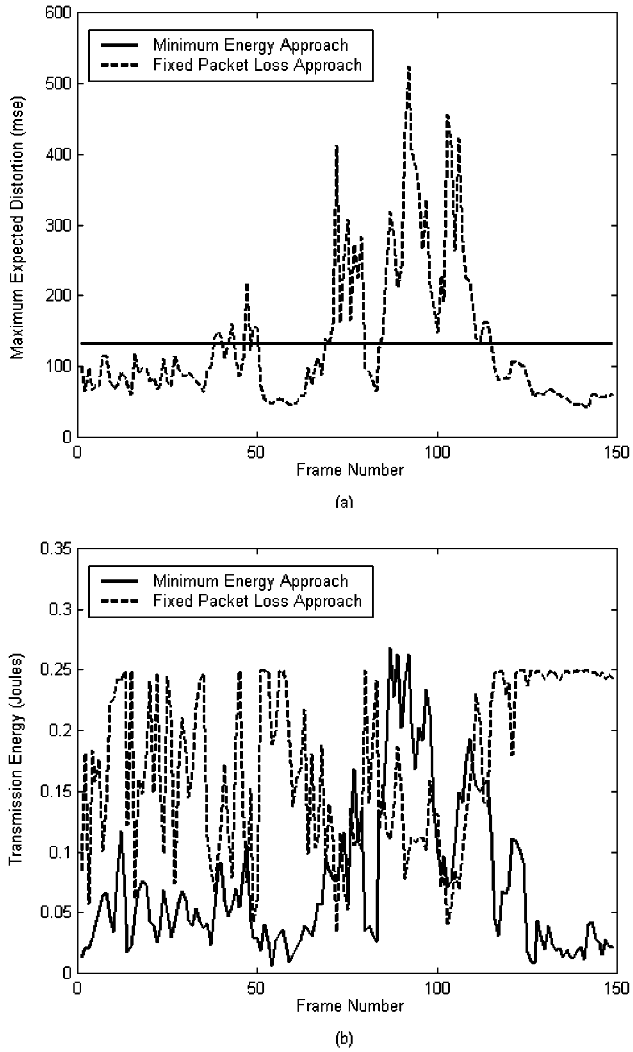

Fig. 9. "Foreman" sequence. (a) Maximum expected distortion per frame. (b) Total transmission energy per frame.

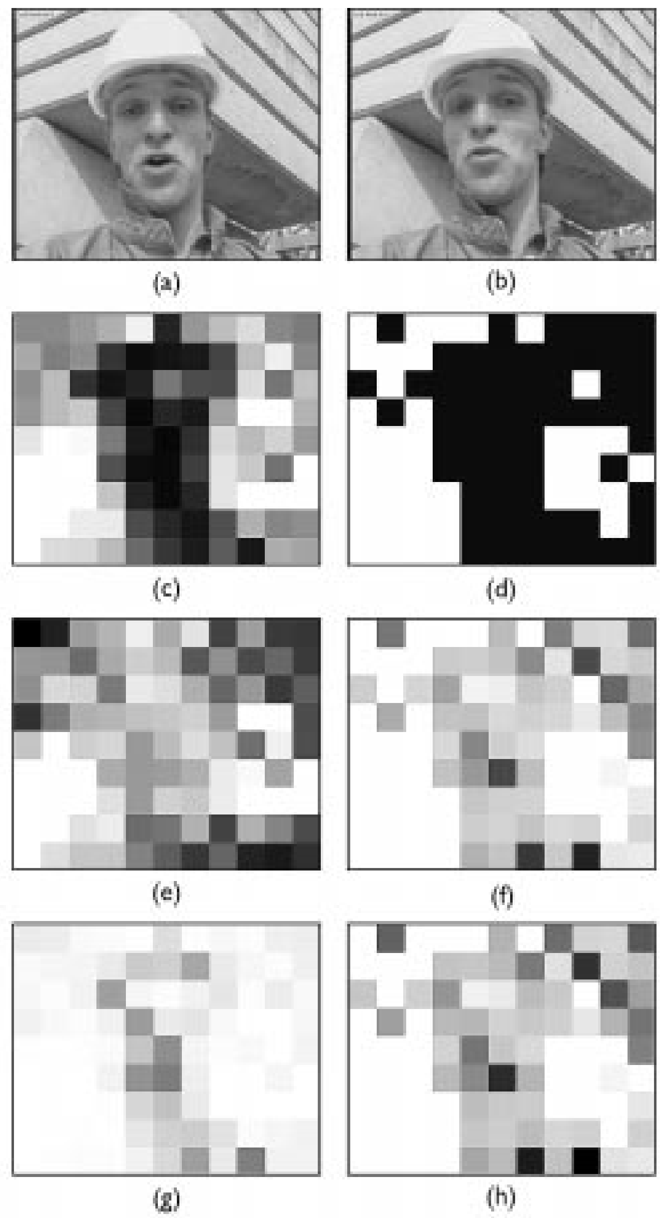

Fig. 10. (a) Frames 42 and (b) Frame 43 in the original "Foreman" sequence. Probability of packet loss per macroblock for frame 43 using the (c) ME approach and (d) FPL approach. Darker macroblocks correspond to a lower probability of packet loss. macroblocks that are not transmitted are shown in white. Bits per macroblock using the (e) ME approach and (f) FPL approach. Darker macroblocks correspond to more bits. Transmission Energy per macroblock using the (g) ME approach and (h) FPL approach. Darker macroblocks correspond to more transmission energy.

packets in a single frame. Frames 41 and 42 of the "Foreman" sequence are shown in Fig. 10(a) and (b), respectively. As shown, the orientation of the foreman's head has changed between the two frames. For frame 42, the ME approach achieves a maximum expected distortion of 132 mse and the FPL approach has a distortion of 131 mse. The total transmission energy for this frame using the ME approach is 0.0469 (Joules) and for the FPL approach it is 0.1155 (Joules). Therefore, both approaches achieve the same expected video quality, but the ME approach uses nearly 60% less energy to transmit this frame.

Recall that the expected distortion depends on both the expected quality if a packet is received, $E[D_R^k]$, and if it is lost, $E[D_L^k]$. In the high activity regions of a frame, $E[D_L^k]$ may be much larger than in regions that have not changed much from the previous frame. Using the ME approach, more protection can be given to the high activity regions in order to increase the likelihood that they will arrive at the decoder correctly. Fig. 10(c) shows the probability of loss for each packet in frame 42 for the ME approach. Darker macroblocks correspond to a smaller probability of packet loss. Macroblocks that are not transmitted are shown in white. As seen in Fig. 10(c), more protection is given to the region of the frame that corresponds to the foreman's head. Therefore, more power is used to transmit this region as opposed to the background. This is because the central region has changed the most significantly between the two frames, and the expected distortion if this region is lost is greater than if the background is lost.

In the comparison approach, the probability of loss is fixed. Therefore, high activity regions are given the same level of protection as the background. Fig. 10(d) shows the probability of loss for each macroblock in frame 42 for the FPL approach. Since the probability of packet loss is fixed, the power used to transmit the region corresponding to the foreman's head is the same as the power used to transmit the background. Because the background has not changed significantly, the expected distortion if a macroblock in the background is lost is small. Thus, in order to achieve the same expected quality, macroblocks in the background do not require the same protection as macroblocks in high activity regions. Therefore, the FPL approach wastes energy by transmitting macroblocks in the background with the same power as macroblocks in the high activity region.

In the FPL approach, the video encoder may allocate more bits to packets in high activity regions, as shown in Fig. 10(f). This is done in order to decrease the expected distortion if those packets are received, $E[D_R^k]$. By decreasing $E[D_R^k]$, the expected distortion for those packets, $E[D^k]$, is also reduced. Because the transmission power is fixed in this approach, more energy is used to transmit packets with more bits, as shown in Fig. 10(h). Therefore, in the FPL approach, more energy may be allocated to high activity regions, but the likelihood of these regions being correctly received is the same as the background. In the ME approach, the bit allocation and power allocation is done jointly. As discussed above, more power can be allocated to regions whose distortion if they are lost, is high. Because energy is a function of the number of bits used to encode a packet and the power used to transmit it, the ME approach finds the optimal bit allocation and protection to give to each packet that uses the least amount of energy. As shown in Fig. 10(g), the ME approach is able to allocate more energy and protection to the regions in a frame that have changed the most significantly. Thus, the ME approach is more energy efficient than considering bit allocation and transmission power management independently.

In order to maintain a constant expected quality across the entire video sequence, our approach is able to conserve energy for periods of high activity. Therefore, the energy saved in transmitting frames in lower activity periods, such as frame 42, can be used for high activity frames. Fig. 11(a) shows frame 92 of the original "Foreman" sequence. This frame is located in the middle of the scene change described earlier. For this frame, the ME approach uses a large amount of transmission energy (0.2631 Joules), in order to achieve a maximum expected distortion of 132 mse. On the other hand, the FPL approach cannot reduce the maximum expected distortion in this frame below 524 mse.

The FPL approach only uses 6006 bits out of a bit budget of 15 kbits. At first glance, this may appear to be counter intuitive. One might think that since this is a frame with high activity, the FPL approach would try to use all the available bits in order to minimize the maximum expected distortion. In fact, that is exactly what the FPL approach does. The macroblocks with the largest $E[D_L^k]$ are encoded using the finest quantizer available. These macroblocks are therefore allocated the maximum number of bits in order to reduce their $E[D_R^k]$. This in turn reduces the expected distortion for these macroblocks, $E[D^k]$. The problem is that since the expected distortion if these packets are lost is so high, reducing the expected distortion if they are received is not enough to decrease their expected distortion. In order to further reduce their expected distortion, more power and protection are needed for these macroblocks. This is something that cannot be achieved if the source coding and transmission power management are done independently. Because the video encoder has already done everything it can to minimize the maximum expected distortion, the remaining macroblocks that have smaller $E[D_L^k]$ require only a few bits in order to achieve the same maximum expected distortion. This is why the FPL approach uses less than the maximum number of bits and energy for this frame. In a sense, the FPL approach is doomed to provide poor video quality when there is a large amount of activity in the video sequence. As shown in Fig. 11(b) and (c), the expected frame at the decoder resembles the original frame much more closely using the ME approach than using the FPL approach.

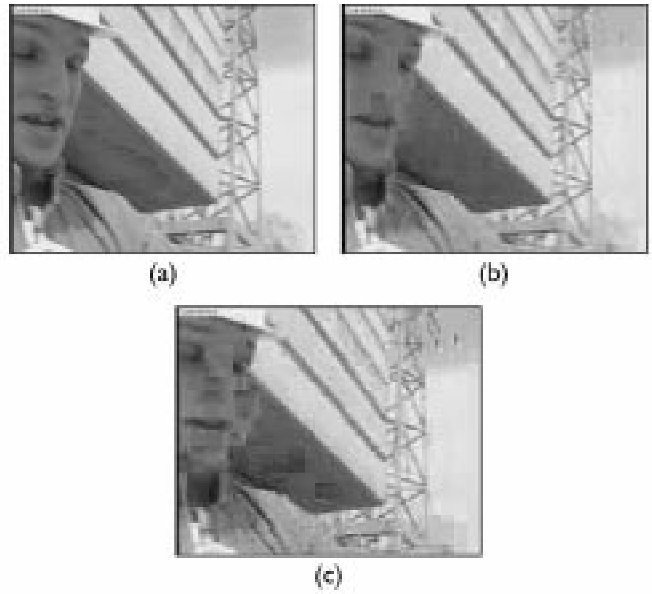

Fig. 11. Frame 92 in the "Foreman" sequence: (a) original frame; (b) expected frame at the decoder using the ME approach; and (c) expected frame at the decoder using the FPL approach.

VII. Conclusions

A method for efficiently utilizing transmission energy in wireless video communications was presented. The objective was to minimize the transmission energy needed to meet both distortion and delay constraints specified by the video application. Source coding and transmission power management were considered jointly. Knowledge of the concealment strategy used by the decoder, as well as the relationship between transmission power and the probability of packet loss, were used to efficiently encode and transmit the video sequence. An algorithm was presented that reduces the computational complexity in finding the minimal energy source coding and power allocation.

Experimental results show that it is more energy efficient to simultaneously adjust the source coding and the transmission power. In order to achieve the same video quality, our approach uses significantly less energy than an approach that considers these factors separately. Similarly, our technique provides higher quality video for the same amount of energy. Our approach provides a method for adaptively allocating resources to different video segments based on their relative importance. Using our approach to wireless video communications, more transmission energy is used during periods of high activity. In addition, our technique allocates more energy and protection to the parts of a frame that have changed most significantly. Therefore, if the background is relatively still, the region of interest will have a higher likelihood of arriving correctly at the decoder than the background. This is because the transmitter knows that the decoder can conceal a missing packet in the background more effectively than a missing packet in a high-activity region. Therefore, our approach provides a method for adaptively allocating resources to different video segments based on their significance.

REFERENCES

[1] N. Bambos, "Toward power-sensitive network architectures in wireless communications: Concepts, issues, and design aspects," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 966–976, June 2000.

[2] M. Zorzi, Ed., *IEEE Personal Commun.*, June 1998, vol. 5.

[3] Y. Wang, G. Wen, S. Wenger, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Processing Mag.*, vol. 17, pp. 61–82, July 2000.

[4] Q. Zhu and Y. Wang, "Error concealment in visual communications," in *Compressed Video Over Networks*, A. R. Reibman and M. T. Sun, Eds. New York: Marcel Dekker, 2000.

[5] A. K. Katsaggelos and N. P. Galatsanos, Eds., *Signal Recovery Techniques for Image and Video Compression and Transmission*. Norwell, MA: Kluwer, 1998, vol. 8, pp. 235–268.

[6] W. M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in *Proc. ICASSP*, vol. 15, Minneapolis, MN, Apr. 1993, pp. 417–420.

[7] M. C. Hong, L. Kondi, H. Scwab, and A. K. Katsaggelos, "Video error concealment techniques," *Signal Processing: Image Commun.*, vol. 14, no. 6–8, pp. 437–492, 1999.

[8] R. O. Hinds, "Robust mode selection for block-motion-compensated video encoding," Ph.D. dissertation, MIT, Cambridge, MA, 1999.

[9] R. O. Hinds, T. N. Pappas, and J. S. Lim, "Joint block-based video source/channel coding for packet-switched networks," *Proc. SPIE*, vol. 3309, pp. 124–133, Jan. 1998.

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

[11] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 952–965, June 2000.

[12] S. Nanda, K. Balachandran, and S. Kumar, "Adaptation techniques in wireless packet data services," *IEEE Commun. Mag.*, vol. 38, pp. 54–64, 2000.

[13] N. Bambos and J. M. Rulnick, "Mobile power management for wireless communication networks," *Wireless Networks*, vol. 3, no. 1, pp. 3–14, 1997.

[14] B. E. Collins and R. L. Cruz, "Transmission policies for time varying channels with average delay constraints," in *Proc. Allerton Conf.*, 1999.

[15] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," IEEE Trans. Inform. Theory, May 2002, to be published.

[16] R. Kravets and P. Krishnan, "Application-driven power management for mobile communication," *Wireless Networks*, vol. 6, pp. 263–277, 2000.

[17] C. Luna, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Transmission energy minimization in wireless video streaming applications," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 4–7, 2001.

[18] B. Prabhakar, E. Uysal-Biyikoglu, and A. El Gamal, "Energy-efficient transmission over a wireless link via lazy packet scheduling," in *Proc. IEEE Infocom*, 2001.

[19] Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Minimizing transmission energy in wireless video communications," in *Proc. ICIP*, Thessaloniki, Greece, 2001, pp. 958–961.

[20] Q. Zhang, W. Zhu, and Y. Zhang, "Network-adaptive scalable video streaming over 3G wireless network," in *Proc. ICIP*, Thessaloniki, Greece, 2001, pp. 579–582.

[21] C. E. Luna, Y. Eisenberg, R. Berry, T. Pappas, and A. K. Katsaggelos, "Joint source coding and data rate adaptation for energy efficient wireless video streaming," in *Proc. Int. Packetvideo Workshop*, Pittsburg, PA, Apr. 2002.

[22] *Video coding for low bit rate communication*, ITU-T Recommendation H.263, Jan. 1998.

[23] "Information technology—Generic coding of audio-visual objects,", ISO/IEC 14 496-2, July 1999.

[24] R. O. Hinds and T. N. Pappas, "Effect of concealment techniques on perceived video quality," in *Proc. SPIE—Human Vision and Electronic Imaging IV*, B. E. Rogowitz and T. N. Pappas, Eds. San Jose, CA, Jan. 1999, vol. 3644, pp. 207–217.

[25] T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing*, A. C. Bovik, Ed. New York: Academic, 2000.

[26] J. Chen and T. N. Pappas, "Perceptual coders and perceptual metrics," in *Proc. SPIE—Human Vision and Electronic Imaging VI*, B. E. Rogowitz and T. N. Pappas, Eds. San Jose, CA, Jan. 2001, vol. 4299, pp. 150–162.

[27] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packet media," IEEE Trans. Multimedia, submitted for publication.

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

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

[30] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Mag.*, vol. 15, pp. 23–50, Nov. 1998.

[31] L. Ozarow, S. Shamai, and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.

Yiftach Eisenberg (S'02) received the B.S. degree in electrical engineering from the University of Illinois at Urbana–Champaign in 1999 and the M.S. degree in electrical engineering from Northwestern University, Evanston, IL, in 2001, where he is currently working toward the Ph.D. degree in electrical engineering.

His current research interests include video processing, error resilience and concealment, and energy-efficient wireless video communications. He is exploring how video quality is related to energy consumption in wireless video communications.

Carlos E. Luna (S'02) received the B.S. degree in electrical engineering from The Johns Hopkins University, Baltimore, MD, in 1993, and the M.S. degree in electrical engineering and the Ph.D. degree in electrical and computer engineering, both from Northwestern University, Evanston, IL, in 1996, and 2002, respectively.

His current research interests include video transmission over packet networks, video compression, and transmission power management for wireless networks.

Thrasyvoulos N. Pappas (M'87–SM'95) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, in 1979, 1982, and 1987, respectively.

From 1987 to 1999, he was a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ. In 1999, he joined the Department of Electrical and Computer Engineering at Northwestern University, Evanston, IL, as an Associate Professor. His research interests are in image and multidimensional signal processing. His recent work has included perceptual image coding, video transmission over lossy channels, model-based halftoning, image and video segmentation, and audiovisual signal processing.

Dr. Pappas is the Electronic Abstracts Editor and an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING. He is the Chair of the IEEE Image and Multidimensional Signal Processing Technical Committee and a member of the Multimedia Signal Processing Technical Committee. He served as Technical Program Co-Chair for ICIP '2001 in Thessaloniki, Greece. He is also Co-Chair for the Conference on Human Vision and Electronic Imaging, sponsored by SPIE and IS&T.

Randall Berry (S'93–M'00) received the B.S. degree in electrical engineering from the University of Missouri-Rolla in 1993 and the M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 1996 and 2000 respectively.

He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at Northwestern University. In 1998, he was a Member of the Technical Staff at MIT Lincoln Laboratory in the Advanced Networks Group. His primary research interests include wireless communication, data networks, and information theory.

Aggelos K. Katsaggelos (F'98) received the Diploma degree in electrical and mechanical engineering from 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, Northwestern University, Evanston, IL, where he is currently a Professor, holding the Ameritech Chair of Information Technology. He is also the Director of the Motorola Center for Telecommunications, and a member of the Associate Staff in the Department of Medicine, Evanston Hospital, Evanston, IL. During the 1986–1987 academic year, he was an Assistant Professor at the Department of Electrical Engineering and Computer Science, Polytechnic University, Brooklyn, NY. He is the Editor of *Digital Image Restoration* (Heidelberg, 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). His current research interests include signal recovery and compression, and multimedia signal processing and communications. He is the co-inventor of eight international patents.

Dr. Katsaggelos has served as a member of the IEEE Signal Processing Society Board of Governors (1999–2001), 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), and a member of the IEEE Technical Committee on Image and Multi-Dimensional Signal Processing (1992–1998). He has served as the General Chairman of the 1994 Visual Communications and Image Processing Conference in Chicago, IL, and as Technical Program Co-Chair of the 1998 IEEE International Conference on Image Processing, Chicago, IL, and is a member of the Publication Board of the IEEE Signal Processing Society, the IEEE TAB Magazine Committee, the Publication Board of the PROCEEDINGS OF THE IEEE, the Scientific Board, the Center for Research and Technology Hellas-CE.R.T.H, the IEEE Technical Committees on Visual Signal Processing and Communications and on Multimedia Signal Processing, an Editorial Board Member of Academic Press, Marcel Dekker, Signal Processing Series, Applied Signal Processing, an area editor for *Graphical Models and Image Processing* (1992–1995), the *Computer Journal*, and Editor-in-Chief of the *IEEE Signal Processing Magazine*. He is 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). He is an Ameritech Fellow and a member of SPIE.