Advances in Efficient Resource Allocation for Packet-Based Real-Time Video Transmission

AGGELOS K. KATSAGGELOS, FELLOW, IEEE, YIFTACH EISENBERG, MEMBER, IEEE, FAN ZHAI, MEMBER, IEEE, RANDALL BERRY, MEMBER, IEEE, AND THRASYVOULOS N. PAPPAS, SENIOR MEMBER, IEEE

Multimedia applications involving the transmission of video over communication networks are rapidly increasing in popularity. Such applications can greatly benefit from adapting video coding parameters to network conditions as well as adapting network parameters to better support the application requirements. These two dimensions can both be viewed as allocating source and network resources to improve video quality. In this paper, we highlight recent advances in optimal resource allocation for real-time video communications over unreliable and resource constrained communication channels. More specifically, we focus on point-to-point coding and delivery schemes in which the sequences are encoded on the fly. We present a high-level framework for resource-distortion optimization. The framework can be used for jointly considering factors across network layers, including source coding, channel resource allocation, and error concealment. For example, resources can take the form of transmission energy in a wireless channel, and transmission cost in a DiffServ-based Internet channel. This framework can be used to optimally trade off resource consumption with end-to-end video quality in packet-based video transmission. After giving an overview of this framework, we review recent work in two areas—energy efficient wireless video transmission and resource allocation for Internet-based applications.

Keywords—*Cross-layer design, energy efficient, error resilience, distortion estimation, internet video, wireless video.*

I. Introduction

Recent years have witnessed a dramatic growth in network-based video applications including on-demand video streaming and distance learning. In the near future, other applications such as video telephony are expected to increase significantly. These applications require transmitting video over communication channels, such as the Internet or wireless channels, that can exhibit wide variability in throughput, delay, and packet loss. Providing acceptable video quality in such environments is a demanding task for both the video encoder/decoder as well as the communication and networking infrastructure. In each of these components, a number of parameters may be adapted based on source content and any available feedback from the network. For example, at the source level one parameter may be the quantization step-size, while at the network level a parameter may be the scheduling of packet transmissions.

There is a growing awareness that both the source and network should be jointly considered in a "cross-layer" perspective, e.g., [1]–[3]. In this paper, we present a high-level framework for optimally adapting both source and network parameters in a video transmission system. These parameters affect the quality of the received video sequence as well as the delay before the sequence may be displayed. Additionally, adapting source and network parameters affects the consumption of limited resources, e.g., energy in a mobile device. In the given framework, we discuss research efforts aimed at allocating resources to balance these considerations.

Our primary focus here is on recent advances in real-time point-to-point video transmission. We use the term real-time to indicate that the video is not prerecoded and must be encoded and delivered within a short time after being captured. Another important area of research is the delivery of pre-encoded media. In these applications, video sequences are typically pre-encoded and stored on a media server. There have been significant advances in the design of media servers including retransmission and packetization techniques [4], rate-distortion optimized packet scheduling [5], and receiver driven multicast approaches [6], [7]. In the remainder of this paper, we will focus on real-time video coding and delivery applications and refer the reader to [2] and [5] for more details on advances in the delivery of pre-encoded media.

In addition to the resource allocation approaches highlighted here, there is significant research activity on new video coding paradigms and network protocols for multimedia communication, some of which is reviewed elsewhere

Manuscript received December 20, 2003; revised July 23, 2004. This work was supported in part by the National Science Foundation under Grant CCR-0311838.

A. K. Katsaggelos, Y. Eisenberg, T. N. Pappas, and R. Berry are with the Electrical and Computer Engineering Department, Northwestern University, Evanston, IL 60208 USA (e-mail: aggk@ece.northwestern.edu; yeisenbe@ece.northwestern.edu; fzhai@ece.northwestern.edu; rberry@ece.northwestern.edu).

F. Zhai was with the Electrical and Computer Engineering Department, Northwestern University, Evanston, IL 60208 USA. He is now with Texas Instruments, Dallas, TX 75243 USA (e-mail: fzhai@ti.com).

Digital Object Identifier 10.1109/JPROC.2004.839621

in this issue, including video coding of H.264/MPEG-4 AVC [8], scalable video coding in MPEG-4 [9], and multiple description coding [10]. Such approaches may provide additional dimensions to be incorporated into the resource allocation framework presented here.

In the following, we first give an overview of a generic video communication system. After reviewing the basic system components, in Section III, we discuss models for packet-based communication channels; in particular, we highlight models for wireless communication channels and models for Internet-based applications. A key component in our optimization framework is the metric used for evaluating video quality; in Section IV, we review approaches for distortion estimation and characterization in packet-based video transmission systems. In Section V, we formulate a general resource-distortion constrained optimization problem that takes into account both resource costs and delay considerations. We then discuss three special cases of this problem in more detail. We conclude in Section VI with some comments on future research directions.

II. Packet-Based Video Transmission Overview

We begin by providing a high-level overview of a packet-based video transmission system. Fig. 1 highlights some of the major conceptual components in such a system. The original video signal is first compressed by the *video encoder*. Compression reduces the number of bits used to describe the video sequence by exploiting both temporal and spatial redundancy. The encoded video will be transmitted over a lossy communication channel and must, therefore, be encoded in an error resilient way that minimizes the effects of losses on the decoded video quality. A recent review of resilient video coding techniques can be found in [11].

In this paper, we focus on one of the most widely utilized video coding techniques, block-based motion compensated (BMC) video coding (e.g., H.263 [12] and MPEG-4 [13]). In this approach, each frame is divided into macro-blocks (MBs) that can either be independently encoded (Intra coded) or predictively coded from a reference MB in a previous frame (Inter coded). For Inter coding, a motion vector (MV) specifies the location of the reference MB in the reference frame. Hence, the name "motion compensated" video coding. Temporal prediction offers increased coding efficiency over Intra coding but is susceptible to error propagation. Transform coding followed by quantization and entropy coding complete the BMC coding process. Let S denote the source coding parameters, such as the coding mode (e.g., inter, intra, or skip) and quantization step-size for all the MBs in a frame (or group of frames). The selection of S affects the source bit rate as well as the end-to-end distortion.

The encoded video sequence is then transmitted over a communication channel. This typically involves packetizing the video stream and passing the packets through the appropriate protocol layers (e.g., RTP/UDP/IP). In Fig. 1, this functionality is implemented in the *transmitter* block. We take a high level view of this block in order to emphasize

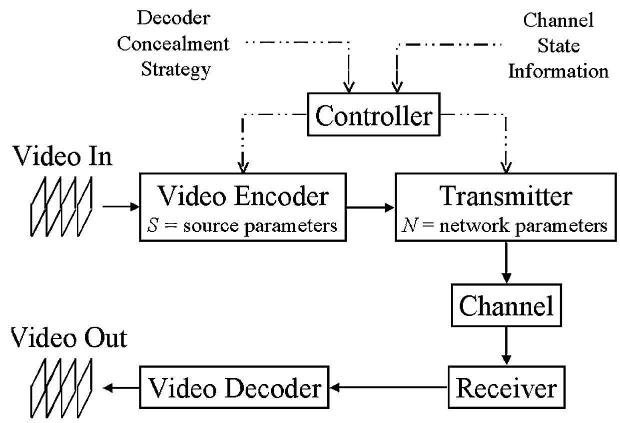


Fig. 1. Packet-based video transmission system diagram.

similarities between a wide range of applications, such as wireless and Internet-based video transmission. Let N represent the set of *network parameters* that can be controlled at the transmitter. This set of parameters depends on the application. For example, in wireless communications, the transmission power and rate can be adapted; in a differentiated services (DiffServ) network, the priority [or quality of service (QoS)] assigned to each packet can be considered a network parameter. In Section III, we discuss other network parameters and models for different communication channels in more detail. In most communication systems, some form of *channel state information* (CSI) is available at the sender, such as an estimate of the fading level in a wireless channel or the congestion over a route in the Internet. This information may be fed back from the receiver and can be used to aid in efficiently allocating resources.[1]

At the *receiver*, the demodulated bit stream is processed by the channel decoder, which performs error detection and/or correction. Corrupt packets can either be passed onto the video decoder or discarded. Here, we assume that only error free information is passed to the video decoder and that corrupt packets are considered lost. This is motivated by the fact that in Internet-based communications, the probability of packet corruption is very low compared to the probability of packet loss caused by congestion in the network. Similarly, in wireless communications, 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 decoder* is responsible for reconstructing the video sequence for display at the receiver. Because some encoded information may have been lost, e.g., due to buffer overflow at a router, the video decoder must conceal any lost information. The commonality among all error concealment strategies is that they exploit redundancy in the received video sequence to conceal lost information. One popular concealment strategy is to use temporal replacement based on the motion information of neighboring macro-blocks [14], [15]. Spatial concealment strategies exploit the correlation between neighboring pixels in the same frame to conceal a lost MB [16]. More complex concealment strategies that use

[1]Another important consideration is the accuracy of any available CSI, which impacts the performance advantage of adaptive approaches.

both temporal and spatial information have also been proposed [17]. A comprehensive review of error concealment techniques can be found in [18].

The *controller* block in Fig. 1 indicates the component of the video transmission system responsible for adapting the source coding parameters S and the network parameters N based on knowledge of the concealment strategy, the source content and any available CSI. As noted in the introduction, we focus on real-time video transmission applications where these parameters are jointly adapted in a cross-layer framework. It is important to point out that the joint controller in Fig. 1 may not be well suited for all packet-based video transmission systems. For example, with streaming media servers, most of the control functionality must be relinquished to the receiver, since the server has to handle hundreds if not thousands of streams simultaneously [2], [6], [7].

The selection of S and N affects the end-to-end distortion $D_{\rm tot}$, the end-to-end delay $T_{\rm tot}$, and the total cost $C_{\rm tot}$ for delivering the video sequence to the end-user. We will use $D_{\rm tot}(S,N), C_{\rm tot}(S,N)$, and $T_{\rm tot}(S,N)$ to explicitly indicate these dependencies. Distortion is caused by both source coding artifacts and channel errors, and will be discussed in greater detail in Section IV. The cost $C_{\rm tot}$ is a measure of the limited resources consumed in transmitting the video sequence, and will be discussed more in Section III. The end-to-end delay $T_{\rm tot}$ is the time between when a video frame is captured at the transmitter and when it is displayed at the receiver [19]. $T_{\rm tot}$ depends in part on the number of bits used to encode the sequence, the transmission rate and any scheduling decisions made by the transmitter.

To better understand the engineering tradeoffs involved in source-network resource adaptation, consider the relationships between $D_{\rm tot}, C_{\rm tot}$, and $T_{\rm tot}$, as shown in Fig. 2. This figure shows the minimum achievable distortion as a function of the total cost and delay. For a given delay $T_{\rm tot}, D_{\rm tot}$ decreases as $C_{\rm tot}$ increases because using more network resources can reduce the probability of packet loss and hence the distortion due to channel errors. The transmission cost required to achieve a certain level of distortion $D_{\rm tot}$ typically decreases with increasing delay $T_{\rm tot}$, as shown in Fig. 2. For example, in a wireless system, the transmission energy required to maintain a fixed probability of error can be reduced by increasing the transmission time and decreasing the transmission power [20], [21]. This observation is used in [22] to provide energy efficient packet transmission over wireless links. Finally, for a given cost budget $C_{\rm tot}, D_{\rm tot}$ increases as $T_{\rm tot}$ decreases. One reason for this is that as $T_{\rm tot}$ decreases, fewer bits may be transmitted across the channel, thus increasing the distortion due to source coding. In Section V, we present a framework for jointly optimizing the network and source coding parameters to manage these tradeoffs for a wide range of applications.

III. Packet-Based Channels

In this section, we take a closer look at the network parameters, N, that can be allocated for each video packet. We also discuss models for how these parameters affect the properties of the communication channel as seen by the video encoder. For video applications, two fundamental properties of the communication channel are the probability of packet loss and the delay needed for each packet to reach the destination. As noted previously, there may also be one or more *costs* associated with the network parameters. The exact nature of these parameters and the costs involved will vary depending on the underlying communication medium and the network protocols that are used. We consider two cases in more detail, namely, Internet and wireless channels.

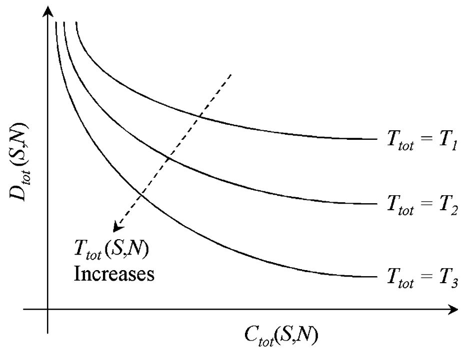


Fig. 2. Relationship between the minimum achievable distortion $D_{\rm tot}(S,N)$, the total cost $C_{\rm tot}(S,N)$, and the end-to-end delay $T_{\rm tot}(S,N)$.

A. Internet Channels

We now discuss models for resource allocation over wire-line networks, and in particular over the Internet.[2] A key difference between a wire-line network and a wireless channel is that losses are much more likely to occur due to buffer overflows and excessive delays as opposed to channel errors and fading. As in the wireless channel, one set of network parameters that can be specified here is the choice of forward error correction (FEC). In the Internet, FEC can be accomplished by sending redundant transport packets using an erasure code, such as a Reed–Solomon code or a Tornado code [23]. In this way, when a transport packet is dropped it can be recovered if a sufficient number of the other packets are received. Several variations of FEC for video communication are surveyed in [1]. Another useful literature survey of rate-distortion optimized error control techniques can also be found in [5]. With FEC, the packet-loss probability can be modeled as specifying a loss probability/packet as a function of the FEC choice. The details of this model will depend on how transport packets are formed from the available video packets [24]. Increasing the amount of FEC results in more packets being sent, which will increase the transmission delay.

In addition to FEC, retransmission-based error control may also be used in the form of ARQ (automatic repeat request) protocols. In ARQ protocols, packets that arrive in error are retransmitted based on feedback from the receiver.

[2]Of course, the Internet also includes wireless links; indeed most future wireless systems will likely have an IP-based architecture. Here, we restrict attention to wire-line connections. An important area of research is to consider hybrid networks that contain both wireless and wire-line links.

Such protocols are only useful if the application can tolerate sufficient delay. When ARQ protocols are used, the decision whether to retransmit a packet or send a new one is another network parameter, which also affects the probability of loss as well as the transmission delay. For pre-encoded video, optimizing over this parameter was considered in [5]. Various hybrid error control techniques that combine aspects of FEC and ARQ have also been proposed, e.g., [1], [7], and [25]–[27].

In the Internet, queuing delays experienced within the network can be a significant delay component. These delays depend in part on the traffic generated by other sessions, as well as the scheduling and buffer management algorithms used at the routers. Queuing delays may be modeled as random variables, the statistics of which may be estimated at the transmitter [5], [28]. In a video communication system, a packet that arrives after its intended playback time is typically considered lost and discarded. This can be modeled by including the delay distribution in the probability of loss calculation, i.e.,

$$\rho_k = \rho_{l,k} + (1 - \rho_{l,k}) \Pr(T_k > \tau_k) \quad (1)$$

where ρ_k is the overall probability of loss for the kth packet, $\rho_{l,k}$ denotes the probability that the packet is dropped in the network, T_k denotes the delay of the kth packet, and τ_k is the maximum allowable delay. To keep buffer delays manageable the Internet uses various congestion control techniques, such as TCP. For video traffic, TCP is not appropriate, but a variety of "TCP-friendly" and other rate control protocols have been proposed [1], [29]. From a resource allocation viewpoint, these protocols impose constraints at the transmitter [28].

In the Internet, a variety of different network level approaches for improving QoS have also been developed; such approaches can also provide network parameters that can be adapted. One example is the DiffServ architecture [30]. In DiffServ, packets are classified into different QoS classes and routers within the network treat packets differently based on their class. For example, each class may have a different probability of loss and/or a different delay distribution associated with it. The assignment of packets to QoS classes can be viewed as another set of network parameters [28], [76], [77]. Each class may also have a different cost, as well as a constraint on the amount of traffic that can be assigned to it.

B. Wireless Channels

In an increasing number of applications, video is transmitted to/or from portable wireless devices including cellular phones, laptop computers connected to wireless local area networks (LANs), or cameras in wireless surveillance systems. Compared to their wire-line counterparts, wireless channels exhibit higher bit error rates, typically have smaller bandwidth, and experience multipath fading and shadowing effects. In this setting, one network parameter that can be specified is the transmission power used to send each packet. For a fixed transmission rate, increasing the transmission power will increase the received signal-to-noise ratio (SNR) and result in a smaller probability of packet loss. This can be modeled by letting the packet loss probability, ρ_k, be given by[3]

$$\rho_k = f(P_k, \theta_k) \quad (2)$$

where P_k is the transmission power used for the kth packet, and θ_k represents the available channel state information (e.g., the fading level). In many systems, the transmitter is able to estimate the channel state, e.g., using a pilot signal. The function f could be determined empirically or modeled analytically. For example, in [31], an analytical model based on the notion of outage capacity [32] is used. In this model, a packet is lost whenever the fading realization results in the channel having a capacity less than the transmission rate. Assuming a Rayleigh fading channel, the resulting relationship is given by

$$\rho_k = f(P_k, \theta_k) = 1 - \exp\left(\frac{1}{P_k \gamma(\theta_k)}(2^{R/W} - 1)\right) \quad (3)$$

where R is the transmission rate (in source bits/sec), W is the bandwidth, and $\gamma(\theta_k)$ is the normalized expected SNR given θ_k. Another possibility for f is to use bounds for the bit error rate for a given modulation and coding scheme; for example, in [33] a model based on the error probability for binary phase-shift keying in a Rayleigh fading channel is used. The cost associated with the transmission is the required energy. In mobile devices, energy consumption is a key consideration for extending battery life. The energy needed to send a packet of L bits with transmission power P is given by $E = PL/R$.

In addition to transmission power, a second network parameter is the transmission rate. At a high-level, a variety of adaptive error control approaches can also be viewed as rate adaptation techniques, including variable rate spreading, adaptive modulation and coding, and various ARQ techniques [34]. Furthermore, different rate adaptation techniques will affect the total performance in different ways; for example in the amount of overhead or delay required. Another important issue is whether these techniques are implemented at the application layer or the link layer. In modern wireless systems, such as third-generation (3G) cellular systems, these parameters can be adapted on a fast time-scale (on the order of 5 ms). For a given transmission power, increasing the transmission rate will increase the probability of error but allow more data to be sent within a given time-period. One way to model this is by expressing ρ_k as

$$\rho_k = f(P_k, R_k, \theta_k) \quad (4)$$

where R_k denotes the transmission rate assigned to the kth packet. For example, in [33] a model was considered where the transmission rate was adapted by changing the amount of FEC applied to each packet using a rate-compatible convolution code. The transmission rate also affects the energy

[3]Of course, other parameters such as the packet-size and amount of interleaving also affect the loss probability and amount of overhead needed.

cost for each packet in two ways—it reduces the transmission time and it increases the power required for a given error probability. In addition to the probability of packet loss, the transmission rate affects the transmission delay incurred by each packet. Video packets that experience excessive delays will typically be considered lost. The allowable delay can be modeled by assigning a "bit budget" or a "time budget" to each frame [31]. For streaming applications, this delay constraint can also be modeled by considering the dynamics of the encoder buffer and the playback buffer at the receiver [35].

IV. Distortion Estimation and Characterization

A critical component in any vide transmission application is a metric for evaluating the quality of the received video signal. From the point of view of the transmitter, the distortion at the receiver is a random variable. To illustrate this point, consider the two reconstructed frames shown in Fig. 3. These frames correspond to what would be seen at the receiver given two different channel loss realizations. Clearly, the distortion at the receiver depends on which packets are lost. Thus, to efficiently encode and transmit a video sequence, the metric used to evaluate the end-to-end distortion, $D_{\rm tot}$, must account for the probabilistic nature of the channel.

Recently, there has been considerable research in the area of distortion estimation and characterization for packet-based video transmission systems [36]–[54]. Here, we highlight two general classes of distortion metrics. The first class consists of methods that measure video quality by the expected distortion in the received sequence. The second class consists of metrics whose aim is to produce more smooth quality by accounting for several sources of distortion variation.

A. Expected Distortion Metrics

The most common metric used to evaluate video quality in communication systems is the expected end-to-end distortion, where the expectation is with respect to the probability loss. The expected distortion for the kth packet can be written as

$$E[D_k] = (1-\rho_k)E[D_{R,k}] + (\rho_k)E[D_{L,k}] \qquad (5)$$

where ρ_k is the probability of loss for the kth packet, $E[D_{R,k}]$ is the expected distortion if the packet is *received* correctly and $E[D_{L,k}]$ is the expected distortion if the packet is *lost*. $E[D_{R,k}]$ accounts for the distortion due to source coding as well as error propagation caused by Inter frame coding. $E[D_{L,k}]$ accounts for the distortion due to concealment. Recall that by adapting N we can affect ρ_k. Predictive coding and error concealment both introduce dependencies between packets. Because of these dependencies, the distortion for a given packet is a function of how other packets are encoded as well as their probability of loss. Accounting for these complex dependencies is what makes calculating the expected distortion a challenging problem.

(a) (b)

Fig. 3. Reconstructed Frame. (a) Simulation 1. (b) Simulation 2.

Work on resilient video coding and transmission for packet lossy networks has primarily focused on minimizing the average expected distortion for a frame (or group of frames) [5], [36]–[47]. In these approaches, the end-to-end distortion is defined as $D_{\rm tot} = (1/K)\sum_{k=1}^{K} E[D_k]$, where K is the number of packets in the frame (or group of frames). The most straightforward approach for calculating the average expected distortion is to simulate several packet loss patterns at the encoder and to average the resulting distortions [37]. Although this approach produces a better estimate of the expected distortion as the number of simulations increases, the drawback is that the computational complexity and storage requirements can quickly become impractical.

Methods for accurately calculating the expected distortion have recently been proposed [38]–[40]. The main contribution of these approaches is that they show that under certain conditions, it is possible to accurately compute the expected distortion with finite storage and computational complexity by using per-pixel accurate recursive calculations. In [38], Hind's method is based on recursively calculating the distribution of the reconstructed value for each pixel in a frame. A recursively accurate method for calculating the expected mean absolute difference is presented in [40]. In [39], Zhang *et al.* develop a powerful algorithm called ROPE, which efficiently calculates the expected mean squared error by recursively computing only the first and second moment of each pixel in a frame. In many advanced video coding schemes, e.g., H.26L and MPEG-4, noninteger motion compensation, deblocking filters, and complex concealment strategies introduce cross-correlation between pixels that make ROPE less precise. Recently, there has been work on approximating these cross-correlation terms in order to extend ROPE to more sophisticated coding schemes [41], [42].

Model-based distortion estimation methods have also been proposed and are useful when the computational complexity and storage capacity are limited. In [43], the authors present a recursive distortion estimation algorithm, which only differs slightly from ROPE in that they approximate the distortion due to concealment. A likely subset of the possible loss patterns is use to estimate the expected distortion in [44]. The authors in [45] and [46] consider the problem of predicting the mean-squared error distortion resulting from different packet loss patterns, including both independent and bursty errors. In [47], He *et al.* develop a model for estimating both source and channel distortion based on the Intra refresh rate and the percentage of zeros among the quantized transform coefficients. Another popular distortion estimation method is to consider the reduction in distortion given that a packet and all the packets it depends on are received, as in [5]. This approach works well when the dependencies between packets are clearly defined, e.g., when transmitting pre-encoded video.

B. Variance-Aware Distortion Metrics

In video coding and transmission there are many sources of quality variation. We review several variance-aware distortion metrics which aim to provide more even quality at the receiver. Reducing the spatial variation in quality across a frame has been considered in order to prevent having good quality in parts of a frame and relatively poor quality in others. In [48] and [49], one attempt at producing more even quality was to minimize the maximum distortion within a frame, i.e., $D_{\mathrm{tot}} = \max_k\{E[D_k]\}$, as opposed to the average distortion.

Controlling temporal variations in quality has also been considered. Rate-control, i.e., assigning bandwidth (bits) to the different frames in a sequence, is related to this type of variation [19], [50]. By allocating more bandwidth during periods of high activity and less to frames with little motion, a rate-control scheme can reduce the overall distortion within a specified time window. Similarly, approaches such as [51], [52], and [49] have explored the benefits of limiting large temporal variations in distortion. It should be noted that for applications with loose enough delay constraints, packet interleaving can be used to spread out bursty errors, thus reducing large temporal and spatial variations in quality.

In video transmission applications, another source of quality variation is due to the random nature of channel errors. At the receiver, the end-user sees only one of many possible reconstructed sequences. Therefore, while the expected distortion computed at the transmitter may be reasonable, the actual distortion at the receiver may vary greatly depending on which packets are lost. In [53] and [54], the authors address this problem by using a distortion metric that accounts for both the mean and the variance of the end-to-end distortion, i.e.,

$$D_{\mathrm{tot}} = (1/K)\sum_{k=1}^{K}\{(1-\alpha)E[D_k] + (\alpha)\mathrm{Var}[D_k]\} \quad (6)$$

where $\mathrm{Var}[D_k]$ is the variance of the distortion for the kth packet, and $\alpha \in [0,1]$ defines the relative importance of the mean and variance of the distortion for a given frame. By accounting for the variance of the distortion, the proposed approach in [53] and [54] makes it more likely that what the end-user sees, closely resembles the mean end-to-end distortion calculated at the transmitter.

V. Resource-Distortion Optimization Framework

A. General Formulation and Optimization Tools

In this paper, we consider techniques that efficiently adapt both source and network parameters in order to minimize the end-to-end distortion while using a limited amount of resources and delay. This optimization can expressed as

$$\min_{\{S,N\}} D_{\mathrm{tot}}(S,N); \quad \text{s.t.: } C_{\mathrm{tot}}(S,N) \le C_0 \quad \text{and} \quad T_{\mathrm{tot}}(S,N) \le T_0 \quad (7)$$

where C_0 is the maximum allowable resource consumption, and T_0 is the end-to-end delay constraint. Here, S denotes the set of source coding parameters that must be optimally adapted. Similarly, N represents the network parameters that can be controlled. We will refer to (7) as the minimum distortion (MD) approach because it provides the best possible quality given cost and delay constraints.

A dual formulation to (7), is to minimize the cost required to transmit a video frame (or group of frames) with an acceptable level of distortion and with tolerable delay

$$\min_{\{S,N\}} C_{\mathrm{tot}}(S,N); \quad \text{s.t.: } D_{\mathrm{tot}}(S,N) \le D_0 \quad \text{and} \quad T_{\mathrm{tot}}(S,N) \le T_0 \quad (8)$$

where D_0 is the maximum allowable distortion. We will refer to (8) as the minimum cost (MC) approach, which is useful for applications with stringent quality of service requirements. For example, (8) can be used to maintain a fixed quality level by varying the resource consumption/frame [31], [32].

Approaches, such as (7) and (8), in which source and network parameters are jointly adapted, have only recently been considered. Before this, a divide and conquer approach was typically used in which these parameters were adapted separately. There are several reasons for this, one of which is the fact that the computational complexity of optimally solving (7) and (8) has been considered infeasible until recently. Advances in low cost and high speed computing are enabling researchers to reconsider these more complex problems. A second reason is that the solution to (7) and (8) is not straightforward, and people in the past have resorted to *ad-hoc* solutions and tweaking parameters. Recent advances in optimal joint source-network resource allocation, required "innovative" application of established optimization methods.

The formulations in (7) and (8) are meant to draw attention to the high-level similarities between various resource allocation problems, some of which are discussed later in this section. It is important to note that applying these formulations to different scenarios requires that the intricacies and challenges of each application be addressed. Next, we highlight some optimization tools which are often employed to solve (7) and (8). Lagrange relaxation can be used to convert the constrained problem (7) into an unconstrained problem

$$\min_{\{S,N\}} J_{\mathrm{tot}} = D_{\mathrm{tot}}(S,N) + \lambda_1 \cdot C_{\mathrm{tot}}(S,N) + \lambda_2 \cdot T_{\mathrm{tot}}(S,N) \quad (9)$$

where $\lambda_1 \ge 0$ and $\lambda_2 \ge 0$ are the Lagrange multipliers. By appropriately choosing λ_1 and λ_2, the solution to (7) can be obtained within a convex-hull approximation by solving (9), [55], [56]. Various methods, such as cutting plane or subgradient methods, can be used to search for λ_1 and λ_2. For a given choice of λ_1 and λ_2, several techniques can be used to find the optimal S and N in (9). For example, dynamic programming (DP) can be used to find the optimal source coding and network parameters when the dependencies between packets are limited, e.g., to a small neighborhood [55], [56]. Dependencies are introduced by source coding (e.g., due to differential encoding) and the concealment strategy (e.g., due to temporal concealment based on the MVs of neighboring MBs). Even when these dependencies are limited, the DP problem is difficult to solve because of the large number of source and network parameters that must be controlled. Iterative decent algorithms can also be used to find a local minimum solution to (9), [5], [57], and [66]. In this approach, parameters are iteratively adapted one at a time while holding the others fixed, until the overall cost is no longer decreasing. Similar optimization techniques can be used to solve the dual formulation (8).

In the following sections, three special cases of the general formulation are addressed. First, we consider efficient source coding techniques. Then, we highlight recent advances in energy efficient wireless video communications, followed by advances in efficient Internet-based video communications.

B. *Efficient Source Coding Techniques*

On the source coding side, there has been much work on transmitting video over unreliable networks. This work has focused on various error resilience techniques for minimizing the effects of lost or corrupt packets. The commonality among these works is that they all assume the channel characteristics, such as the probability of packet loss, are given and cannot be controlled. The objective is then to adapt the source coding parameters S in order to minimize the end-to-end distortion while satisfying delay constraints imposed by the application. This objective is clearly a special case of the general resource-distortion formulation (7), and can be written as

$$\min_{\{S\}} D_{\mathrm{tot}}(S); \quad \text{subject to: } T_{\mathrm{tot}}(S) \leq T_0. \qquad (10)$$

Note that for a channel with a fixed transmission rate, the delay constraint in (10) can be converted to an equivalent source rate or bit budget constraint.

Resilient video coding is a well-studied field of research and includes techniques such as mode and quantizer selection, data partitioning [11], reversible variable length coding [58], and packetization [43]. In this section, we review recent advances in mode (and quantizer) selection algorithms. For details on other efficient source coding techniques, we refer the reader to [11], as well as papers found elsewhere in this issue.

Mode selection algorithms minimize the end-to-end distortion by selecting the coding mode and quantization parameters for each MB in the sequence. In other words, given the delay (or equivalently rate) constraint in (10), these techniques tackle the problem of trading off the compression efficiency of Inter coding with the resilience to error propagation of Intra coding. In [38]–[40], and [43], mode selection is based on recursive estimates of the expected distortion. Mode selection based on an error sensitivity metric is used in [59]. Algorithms for selecting the Intra refresh rate, i.e., the number of Intra MBs/frame, has been studied in [60], and more recently in [47] and [61]. In [53] and [54], mode selection is based on the mean and variance of the distortion.

Mode selection algorithms have traditionally focused on single-frame BMC coding (SF-BMC), i.e., they consider the case where the reference for the current frame is defined to be the previous frame. Recently, there has been significant work on mode selection using multiple-frame BMC (MF-BMC). Unlike SF-BMC, these approaches choose the reference frame from a group of previous frames. MF-BMC techniques capitalize on the correlation between multiple frames to improve compression efficiency and increase error resilience. One drawback of MF-BMC is the large computational complexity and storage capacity needed to perform optimal mode selection in this setting. The authors in [37] and [44] consider the possibility of selecting a different reference frame for each MB in the current frame, as allowed in the latest standard H.264/MPEG4 AVC [8]. In [41] and [62], only two possible reference frames, i.e., a short term and a long term reference, are considered as a means to derive the benefits of MF-BMC while limiting storage and complexity requirements. Markov chain analysis is used in [63] to demonstrate the improved robustness of MF-BMC over SF-BMC.

C. *Energy Efficient Wireless Video Communications*

Wireless video communications is a broad, active, and well-studied field of research [64], [65]. In this section, we focus on one aspect of this general area—resource allocation for energy efficient wireless video communications. We consider the case where video is transmitted from a mobile wireless device. Such devices typically rely on a battery with a limited energy supply; efficiently utilizing this energy is, thus, a critical issue.[4]

Recently, several adaptation techniques have been proposed specifically for energy efficient wireless video transmission. A trend in this field of research is the joint adaptation of source coding and transmission parameters (S and N) based on the time-varying source content and channel conditions. The general resource-distortion framework, i.e., (7) and (9), therefore, encompasses these techniques with C_{tot} defined to be the total energy consumed in delivering the video sequence. Here, we focus on communication over a single wireless link.[5]

For video sequences, sensitivity to channel errors varies over time and even across packets in the same frame. Joint source coding and power allocation techniques account for this varying error sensitivity by adapting the transmission power/packet based on the source content. In other words, these techniques use transmission power as an unequal error

[4]Minimizing transmission energy can also reduce interference between users and increase network capacity. Thus, the techniques considered here may also be useful in applications where video is transmitted from a base station to a mobile device.

[5]Study of the single wireless link provides insight into one component of more complicated scenarios involving multiple links.

protection mechanism. Video transmission over code division multiple access networks using a scalable source coder (3-D SPIHT), along with error control and power allocation is considered in [66]. A scheme is presented in [67] for meeting the expected peak signal-to-noise ratio (PSNR) requirements of each user by adapting the bit-error rate. A scheme for allocating source rate and transmission power under bandwidth constraints is considered in [68]. In [31], optimal mode and quantizer selection is considered jointly with transmission power allocation. In [53] and [54], variance-aware resource allocation is proposed.

Joint source coding and transmission rate adaptation has also been studied as a means of providing energy efficient video communications. In order to maintain a certain probability of loss, the energy consumption increases as the transmission rate increases [20], [21]. Therefore, in order to reduce energy consumption it is advantageous to transmit at the lowest rate possible [22]. In addition to affecting energy consumption, the transmission rate determines the number of bits that can be transmitted within a given period of time. Therefore, as the transmission rate decreases, the distortion due to source coding increases. Joint source coding and transmission rate adaptation techniques adapt the source coding parameters and the transmission rate in order to trade off energy consumption with end-to-end video quality. In [35], the authors consider optimal source coding and transmission rate adaptation. Stochastic Dynamic Programming is used to find an optimal source coding and transmission policy based on a Markov state channel model. A key idea in this work is that performance can be improved by allowing the transmitter to suspend or slow down transmissions during periods of poor channel conditions, as long as the delay constraint is met.

In addition to transmission power, processing power is another source of energy consumption, which is particularly important in applications with short-range communication requirements. Recently, there has been work on minimizing the total energy consumption in mobile devices by considering both the transmission and processing power [69], [70]. These techniques utilize complexity adaptive source coders in order to tradeoff improved compression efficiency with processing power. For example, in [69] the authors consider the processing power used in motion estimation. Better motion estimation requires more computations (and, thus, more processing power), but provides improved compression efficiency. In [70], total system energy minimization is considered for wireless image transmission. With respect to reducing processing power, we should mention related work, such as [71] and [72], in which the idea is to shift computational complexity from the encoder to the decoder.

To illustrate some advantages of jointly adapting the source coding and transmission parameters in wireless video transmission systems, we present experimental results which are discussed in detail in [31], [53], and [54]. We compare a joint source coding and transmission power allocation (JSCPA) approach, i.e., (7), with an independent source coding and power allocation (ISCPA) approach in which S and N are independently adapted. In Fig. 4, we plot the expected PSNR/frame of both approaches for the "foreman" test sequence coded at 15 frames/s (fps). It is important to note that both approaches use the same transmission energy and delay/frame. In this experiment, the *generalized skip* option, introduced in [31], was used to improve efficiency. The idea is that if the concealment of a certain packet results in sufficient quality, then the algorithm can intentionally not transmit this packet in order to allocate additional resources to packets that are more difficult to conceal.

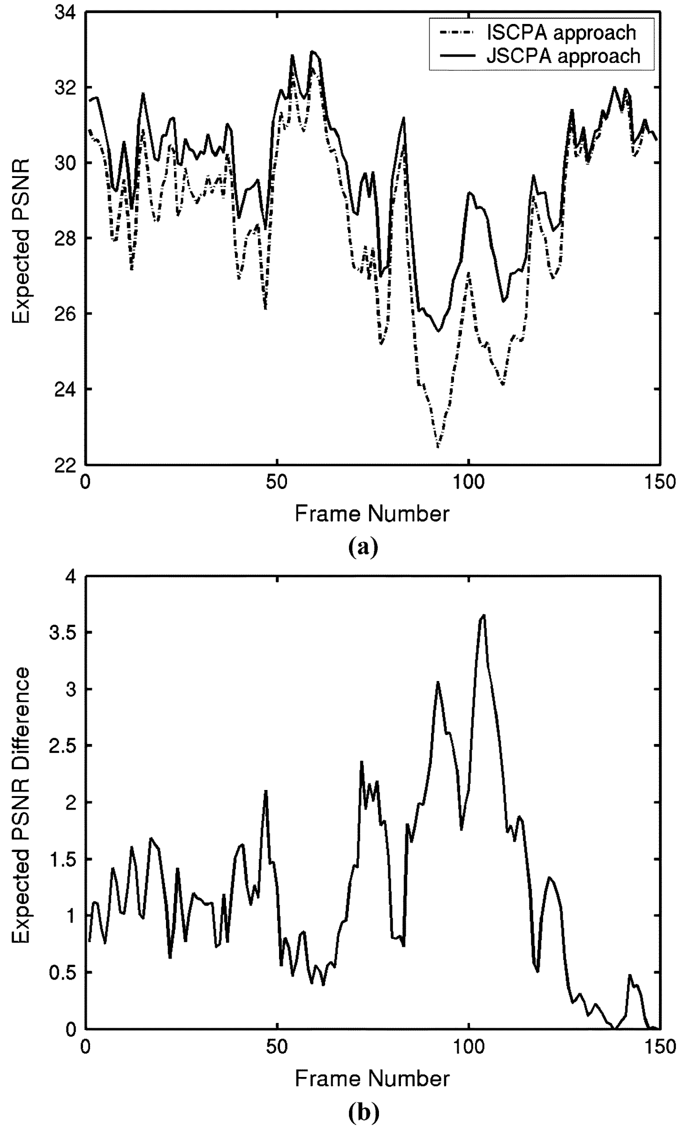


Fig. 4. (a) Expected PSNR/frame for the ISCPA and JSCPA approaches. (b) Difference in expected PSNR between the two approaches.

As shown in Fig. 4, the JSCPA approach achieves significantly higher quality (expected PSNR)/frame than the ISCPA. Because the video encoder and the transmitter operate independently in the ISCPA approach, 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 and adapts the power in order to maintain a constant probability of packet loss. The JSCPA on the other hand is able to adapt the power/packet and, thus, the probability of loss, based on the relative importance of each packet. For example, more power can be allocated to packets that are difficult to conceal. As shown in Fig. 4, the PSNR improvement is greatest during periods of high activity. For example, around frame 100 there is a scene change in which the camera pans from the foreman to the construction site. During this time, the JSCPA approach achieves PSNR improvements of up to 3.5 dB. This gain comes from the ability of the JSCPA approach to increase the power while decreasing the number of bits sent in order to improve the reliability of the transmission. The ISCPA scheme is unable to adapt the protection level and, thus, incurs large distortion during periods of high source activity.

D. *Efficient Internet-Based Video Communications*

Like wireless video transmission, Internet-based video communications is an active and extensively studied field of research, e.g., see [1]. In this section, we highlight two classes of efficient resource allocation techniques for Internet-based video communications. The first is joint source channel coding (JSCC) techniques. The second is techniques for video transmission over DiffServ networks.

1) Joint Source Channel Coding (JSCC): Various JSCC approaches for video or image transmission have been widely studied, e.g., [1], [24], [36], [47]. Here, we focus on approaches for Internet-based applications. As the name suggests, JSCC[6] approaches jointly adapt the source and channel coding parameters in order adapt to the time-varying source content and channel conditions. Their objective is to minimize the end-to-end distortion by efficiently allocating the available bandwidth (bit rate) between source and channel coding [1]. Therefore, JSCC is a special case on the general resource-distortion framework presented in Section IV-A, and can be written as

$$\min_{\{S,N\}} D_{\mathrm{tot}}(S,N); \mathrm{s.t.}: T_{\mathrm{tot}}(S,N) \leq T_0. \tag{11}$$

As discussed in Section III, FEC and ARQ are two basic error correction techniques. Of the two, FEC has been widely suggested for real-time video applications due to the strict delay requirement and semi-reliable nature of video streams [1], [11], [73], [74]. However, ARQ may have several advantages depending on the application. One disadvantage of FEC is that it incurs constant transmission overhead even when the channel is loss free. ARQ on the other hand can automatically adapt to the channel loss characteristics by retransmitting only lost packets. Thus, if the application has a relatively loose end-to-end delay constraint (e.g., on-demand video streaming applications), retransmission may be more applicable. Even for real-time applications, delay constrained application-layer ARQ has been considered and shown to be useful for video streaming in certain situations [25].

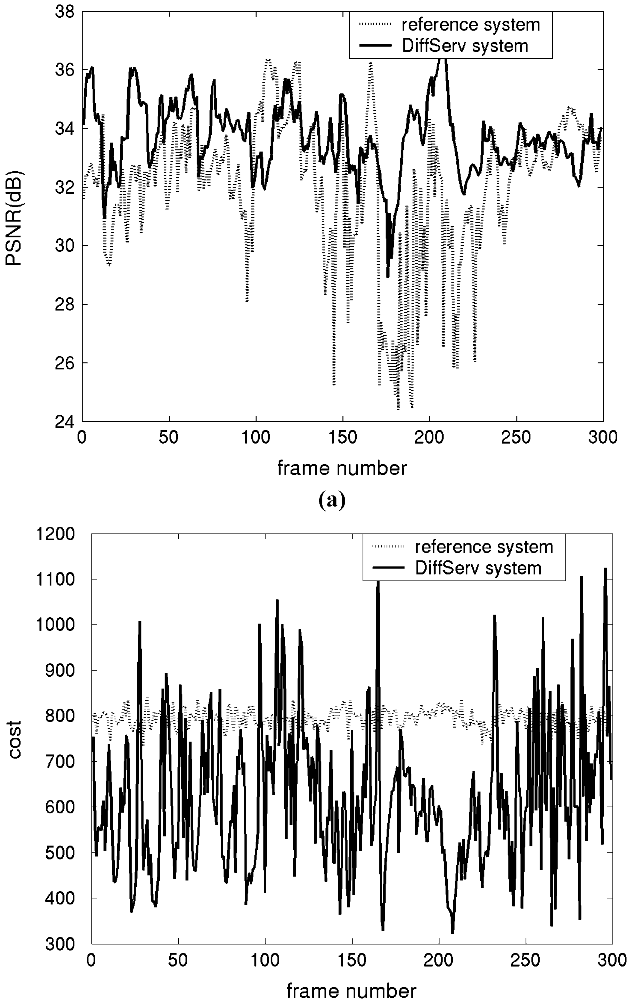
In terms of hybrid FEC/retransmission, for wireless IP networks, a link-layer hybrid FEC/ARQ scheme is considered in [26], and a heuristic application-layer hybrid FEC/ARQ technique is proposed for video transmission in [25]. A receiver-driven hybrid FEC/Pseudo-ARQ mechanism is proposed for Internet multimedia multicast in [7]. In [27], optimal error control is performed by jointly considering source coding with hybrid FEC and sender-driven application-layer selective retransmission. This study is carried out in the rate-distortion optimization framework, as in (11), with a sliding window scheme in which lost packets are selectively retransmitted. Simulations in [27] show that the performance advantage of using either FEC or selective retransmission depends on the packet loss rate and the round-trip time. In that work, the proposed hybrid FEC and selective retransmission approach is able to derive the benefits of both approaches by adapting the type of error control to the channel conditions.

2) Differentiated Services Networks (DiffServ): Today's Internet is a best effort network and, thus, does not provide any guaranteed QoS. As discussed in Section III, DiffServ has recently been considered as a means of providing discriminate quality of service based on service classes. In this architecture, each class is associated with a price/transmitted bit, byte, or packet. Typically, transmitting a packet in a higher priority service class results in a higher cost but a better QoS (lower delay and loss probability).

In [75], an adaptive packet forwarding mechanism was proposed for a DiffServ network where video packets are mapped onto different DiffServ service levels. The authors in [76] proposed a rate-distortion optimized packet marking technique to deliver MPEG2 video sequences in a DiffServ IP network. Their goal was to minimize the bandwidth consumption in the premium class while achieving nearly constant perceptual quality. In [77], cost-distortion optimized multimedia streaming over DiffServ networks was studied for pre-encoded media. However, the work in [75]–[77] does not incorporate video source coding.

A formulation as in (7) for transmitting video over a DiffServ network is studied in [28] and [78]. In this paper, the network parameters are the QoS class and the scheduling decision for each packet. To illustrate the advantage of jointly selecting the source coding parameters and QoS class, we consider a reference system similar to today's Internet where only one QoS class is available. In this reference system, source coding decisions are made to minimize the expected end-to-end distortion subject to the transmission delay constraint. The corresponding DiffServ system matches the delay and cost of the reference system. Simulations based on the parameter settings given in [28] showed that the proposed joint coding and packet marking approach outperforms the corresponding single-class reference system. Fig. 5 shows the performance comparison for the "foreman" test sequence at 30 fps. Note that the results in Fig. 5 are for a real-time system with tight delay constraints. When the objective is to minimize the end-to-end distortion, i.e., (7), the proposed DiffServ approach achieves an average PSNR improvement of 1.3 dB over the single-class reference system. Similarly, for the MC formulation, (8), an average cost savings of 20%/frame is achieved using the proposed technique.

[6]Note that the term JSCC has several interpretations, e.g., JSCC also describe schemes that use a single code book for source and channel coding.



(a)

(b)

Fig. 5. Comparison of the DiffServ approach with the single-class reference system. (a) MD approach, (b) MC approach.

VI. Future Research Directions

In this paper, we have reviewed recent advances in efficient resource allocation techniques for packet-based video transmission systems. In the future, there are a number of important research issues in this area that deserve significant attention. Although we have focused on single layer block-based motion compensated (BMC) video coding, the resource-distortion framework presented here can be applied to other techniques, such as scalable video coding. For example, advances in MPEG4 fine granularity scalability (FGS) and motion compensated wavelet video coding have recently made the compression efficiency of these techniques competitive with BMC coding. Utilizing scalable coding schemes within a resource-distortion framework is a promising research direction. In this setting, adapting the network parameters based on the source content may be simpler due to the prioritized representation of data in scalable coding.

Cross-layer design of heterogeneous wireless Internet video systems is an active field of research that warrants further investigation. Current work includes deploying a proxy at the edge between the wired and wireless network to improve video quality [79]. Other work has focused on end-to-end error control methods for dealing with the unique characteristics of each network component [80], [81].

The techniques encompassed by the resource-distortion framework considered here all assume that the constraints, such as cost and delay, are set/frame (or group of frames) based on the application. Developing algorithms for setting these constraints in a smart way is an important research direction. This is similar to rate control, but differs in that these schemes must effectively distribute both bandwidth (delay) and resources, such as energy, over time. Also, when looking over longer time-horizons, tractable approaches that balance efficiency with complexity will be needed.

This paper primarily focuses on efficient resource allocation for a single user in a unicast scenario. Efficient resource allocation for multiuser and multicast video transmission systems is another active research area with unique challenges. Issues such as fairness as well as distributed versus centralized resource allocation must be addressed in this setting.

Extending the resource-distortion framework to new coding paradigms, e.g., distributed signal processing and sensor networks, is another promising research direction. Here, the definition of cost and distortion must be modified appropriately based on the application. For example, to ensure network connectivity, energy constraints may be based on the remaining battery life of each node in the network.

Finally, one of the most critical aspects of designing efficient resource allocation algorithms is an accurate metric for evaluating the end-to-end video quality. An area of research we feel requires significant work is the development of perceptually motivated objective quality metrics for packet-based video transmission. By gaining a better understanding of how the adaptation of certain source and network parameters affects the perceived video quality, algorithms can be designed to more effectively allocate limited resources.

Acknowledgment

The authors would like to acknowledge Dr. C. Luna for valuable discussions related to this paper.

References

[1] D. Wu, Y. T. Hou, and Y.-Q. Zhang, "Transporting real-time video over the Internet: Challenges and approaches," *Proc. IEEE*, vol. 88, no. 12, pp. 1855–1877, Dec. 2000.

[2] B. Girod, J. Chakareski, M. Kalman, Y. J. Lang, E. Setton, and R. Zhang, "Advances in network-adaptive video streaming," presented at the Tyrrhenian Int. Workshop Digital Communications, Capri, Italy, Sept. 2002, pp. 1–8.

[3] H. Zheng, "Optimizing wireless multimedia transmission through cross layer design," in *Proc. ICME*, Baltimore, MD, July 2003, pp. 185–188.

[4] G. J. Conklin, G. S. Greenbaum, K. O. Lillevold, A. F. Lippman, and Y. A. Reznik, "Video coding for streaming media delivery on the Internet," *IEEE Trans. Circuits Syst, Video Technol.*, vol. 11, no. 3, pp. 269–281, Mar. 2001.

[5] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," , 2001, submitted for publication.

[6] S. McCanne, M. Vetterli, and V. Jacobson, "Low-complexity video coding for receiver-driven layered multicast," *IEEE J. Sel. Areas Commun.*, vol. 15, no. 6, pp. 983–1001, Aug. 1997.
[7] P. A. Chou, A. E. Mohr, A. Wang, and S. Mehrotra, "Error control for receiver-driven layered multicast of audio and video," *IEEE Trans. Multimedia*, vol. 3, no. 1, pp. 108–122, Mar. 2001.
[8] G. Sullivan and T. Wiegand, "Video compression—From concepts to the H.264/AVC standard," in *Proc. IEEE*, vol. 93, 2005, pp. 18–31.
[9] J. Ohm, "Advances in scalable video coding," in *Proc. IEEE*, vol. 93, 2005, pp. 42–56.
[10] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," in *Proc. IEEE*, vol. 93, 2005, pp. 57–70.
[11] Y. Wang, G. Wen, S. Wenger, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Process. Mag.*, vol. 17, pp. 61–82, July 2000.
[12] Video Coding for Low Bitrate Communications, ITU, ITU Telecom. Standardization Sector of ITU, Draft ITU-T Recommendation H.263 Version 2, Sept. 1997.
[13] *Information Technology—Generic Coding of Audio-Visual Objects*, July 1999. ISO/IEC 14 496-2.
[14] W. M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," *Proc. SPIE Visual Comm. Image Processing*, pp. V417–V420.
[15] M. C. Hong, L. Kondi, H. Scwab, and A. K. Katsaggelos, "Video error concealment techniques," *Signal Process.: Image Communications*, vol. 14, no. 6–8, pp. 437–492, 1999.
[16] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projection onto convex sets," in *IEEE Trans. Image Process.*, Apr. 1995, pp. 470–477.
[17] S. Belfiore, M. Grangetto, E. Magli, and G. Olmo, "Spatio-temporal video error concealment with perceptually optimized mode selection," in *Proc. ICME*, Baltimore, MD, July 2003, pp. 169–172.
[18] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.
[19] C. Y. Hsu, A. Ortega, and M. Khansari, "Rate control for robust video transmission over burst-error wireless channels," *IEEE J. Select. Areas Comm.*, vol. 17, no. 5, pp. 756–773, May 1999.
[20] F. Adler, "Minimum energy cost of an observation," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 28–32, 1955.
[21] R. G. Gallager, "Energy Limited Channels: Coding, Multi-Access, and Spread Spectrum," Massachusetts Insy. Technol., Cambridge, LIDS Rep. LIDS-P-1714, Nov. 1987.
[22] E. Uysal-Biyikoglu, B. Prabhakar, and A. El Gamal, "Energy-efficient packet transmission over a wireless link," *IEEE Trans. Networking*, vol. 10, no. 8, pp. 487–499, Aug. 2002.
[23] M. Luby *et al.*, "Practical loss-resilient codes," presented at the 29th ACM Symp Theory of Computing, Atlanta, GA, 1997.
[24] F. Zhai, Y. Eisenberg, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Packetization schemes for forward error correction in internet video streaming," presented at the 41st Allerton Conf. Communication, Control, and Computing, Monticello, IL, Oct. 2003.
[25] F. Hartanto and H. R. Sirisena, "Hybrid error control mechanism for video transmission in the wireless IP networks," presented at the IEEE 10th Workshop Local and Metropolitan Area Networks (LANMAN'99), Sydney, Australia, Nov. 1999.
[26] S. Falahati, A. Svensson, N. C. Ericsson, and A. Ahlen, "Hybrid type-II ARQ/AMS and scheduling using channel prediction for downlink packet transmission on fading channels," presented at the Nordic Radio Symp., Nanashamn, Sweden, 2001.
[27] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Rate-distortion optimized hybrid error control for real-time packetized video transmission," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, 2004, pp. 1318–1322.
[28] F. Zhai, C. Luna, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and packet classification for real-time video transmission over differentiated services networks," *IEEE Trans. Multimedia*, submitted for publication.
[29] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "Resource allocation for multimedia streaming over the internet," *IEEE Trans. Multimedia*, vol. 3, no. 3, pp. 339–355, Sept. 2001.
[30] S. Blake *et al.*, An Architecture for Differentiated Services, IETF, RFC 2475, Dec. 1998.
[31] Y. Eisenberg, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and transmission power management for energy efficient wireless video communications," *IEEE Trans. Circuits Syst, Video Technol.*, vol. 12, no. 6, pp. 411–424, June 2002.
[32] L. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, no. 2, pp. 359–378, May 1994.
[33] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source-channel coding and power allocation for energy efficient wireless video communications," in *Proc. 41st Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2003.
[34] S. Nanda, K. Balachandran, and S. Kumar, "Adaptation techniques in wireless packet data services," *IEEE Commun. Mag.*, vol. 38, no. 1, pp. 54–64, Jan. 2000.
[35] C. E. Luna, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and data rate adaptation for energy efficient wireless video streaming," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1710–1720, Dec. 2003.
[36] 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.
[37] T. Stockhammer, T. Wiegand, and S. Wenger, "Optimized transmission of H.26L/JVT coded video over packet-lossy networks," in *Proc. ICIP*, Rochester, NY, 2002, pp. 173–176.
[38] R. O. Hinds, "Robust mode selection for block-motion-compensated video encoding," Ph.D. dissertation, Massachusetts Inst. Technol, Cambridge, MA, 1999.
[39] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Sel. Areas Commun.*, no. 6, pp. 966–976, June 2000.
[40] D. Wu, Y. T. Hou, B. Li, W. Zhu, Y.-Q. Zhang, and H. J. Chao, "An end-to-end approach for optimal mode selection in Internet video communication: Theory and application," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 977–995, June 2000.
[41] A. Leontaris and P. C. Cosman, "Video compression with intra/inter mode switching and a dual frame buffer," presented at the IEEE Data Compression Conf., Snowbird, UT, Mar. 25–27, 2003.
[42] H. Yang and K. Rose, "Recursive end-to-end distortion estimation with model-based cross-correlation approximation," presented at the ICIP, Barcelona, Spain, Sept. 2003.
[43] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 952–965, June 2000.
[44] T. Wiegand, N. Farber, K. Stuhlmuller, and B. Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1050–1062, June 2000.
[45] Y. J. Liang, J. Apostolopoulos, and B. Girod, "Analysis of packet loss for compressed video: Does burst-length matter," in *Proc. ICASSP*, vol. 5, Hong Kong, China, 2003, pp. 684–687.
[46] J. Chakareski, J. Apostolopoulos, W. Tan, S. Wee, and B. Girod, "Distortion chains for predicting the video distortion for general packet loss patterns," in *Proc. ICASSP*, QC, Canada, 2004.
[47] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst, Video Technol.*, vol. 12, no. 6, pp. 511–523, June 2002.
[48] G. M. Schuster, G. Melnikov, and A. K. Katsaggelos, "A review of the minimum maximum criterion for optimal bit allocation among dependent quantizers," *IEEE Trans. Multimedia*, vol. 1, no. 1, pp. 3–17, Mar. 1999.
[49] Y. Eisenberg, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Optimal source coding and transmission power management using a min-max expected distortion approach," in *Proc. ICIP*, Rochester, NY, 2002, pp. 537–540.
[50] A. R. Reibman and B. G. Haskell, "Constraints on variable bit-rate video for ATM networks," *IEEE Trans. Circuits Syst, Video Technol.*, vol. 2, no. 4, pp. 361–372, Dec. 1992.
[51] Y. Sermadevi, M. Masry, and S. S. Hemami, "MINMAX rate control with a perceived distortion metric," presented at the SPIE Conf. Visual Communications and Image Process., San Jose, CA, Jan. 2004.
[52] A. Munteanu, Y. Andreopoulos, M. van der Schaar, P. Schelkens, and J. Cornelis, "Control of the distortion variation in video coding systems based on motion compensated temporal filtering," presented at the ICIP, Barcelona, Spain, Sept. 2003.
[53] Y. Eisenberg, F. Zhai, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Variance-aware distortion estimation for wireless video communications," presented at the ICIP, Barcelona, Spain, Sept. 2003.

[54] ——, "VAPOR: Variance-aware per-pixel optimal resource allocation," *IEEE Trans. Image Process.*, submitted for publication.

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

[56] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 23–50, Nov. 1998.

[57] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. New York: Wiley, 1987.

[58] J. Wen and J. Villasenor, "Reversible variable length codes for robust image and video transmission," presented at the Asilomar, Pacific Grove, CA, Nov. 1997.

[59] J. Liao and J. Villasenor, "Adaptive intra update for video coding over noisy channels," presented at the ICIP, Laussane, Switzerland, Oct. 1996.

[60] P. Haskell and D. Messerschmitt, "Resynchronization of motion-compensated video affected by ATM cell loss," in *Proc. ICASSP*, vol. 3, 1992, pp. 545–548.

[61] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1012–1032, June 2000.

[62] T. Fukuhara, K. Asai, and T. Murakami, "Very low bit-rate video coding with block partitioning and adaptive selection of two time-differential frame memories," *IEEE Trans. Circuits Syst, Video Technol.*, vol. 7, no. 1, pp. 212–220, Feb. 1997.

[63] M. Budagavi and J. D. Gibson, "Multiframe video coding for improved performance over wireless channels," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 252–265, Feb. 2001.

[64] B. Girod and N. Farber, "Wireless video," in *Compressed Video Over Networks*, M.-T. Sun and A. R. Reibman, Eds. New York: Marcell-Dekker, 2001, pp. 465–511.

[65] D. Wu, Y. T. Hou, and Y.-Q. Zhang, "Scalable video coding and transport over broad-band wireless networks," in *Proc. IEEE*, vol. 89, Jan. 2001, pp. 6–20.

[66] S. Zhao, Z. Xiong, and X. Wang, "Joint error control and power allocation for video transmission over CDMA networks with multiuser detection," *IEEE Trans. Circuits Syst, Video Technol.*, vol. 12, no. 6, pp. 425–437, June 2002.

[67] I.-M. Kim and H.-M. Kim, "A new resource allocation scheme based on a PSNR criterion for wireless video transmission to stationary receivers over gaussian channels," *IEEE Trans. Wireless Comm.*, vol. 1, no. 3, pp. 393–401, July 2002.

[68] Y. S. Chan and J. W. Modestino, "Transport of scalable video over CDMA wireless networks: A joint source coding and power control approach," in *Proc. ICIP*, 2001, pp. 973–976.

[69] Q. Zhang, Z. Ji, W. Zhu, and Y.-Q. Zhang, "Power-minimized bit allocation for video communication over wireless channels," *IEEE Trans. Circuits Syst, Video Technol.*, vol. 12, no. 6, pp. 398–410, June 2002.

[70] S. Appadwedula, M. Goel, N. R. Shanbhag, D. L. Jones, and K. Ramchandran, "Total system energy minimization for wireless image transmission," *J. VLSI Signal. Process.*, pp. 99–117, Feb. 2001.

[71] R. Puri and K. Ramchandran, "PRISM: A "reversed" multimedia coding paradigm," presented at the ICIP, Barcelona, Spain, Sept. 2003.

[72] A. Aaron, E. Setton, and B. Girod, "Toward practical Wyner-Ziv coding of video," presented at the ICIP, Barcelona, Spain, Sept. 2003.

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

[74] R. Zhang, S. L. Regunathan, and K. Rose, "End-to-end distortion estimation for RD-based robust delivery of pre-compressed video," presented at the 35th Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA, Oct. 2001.

[75] J. Shin, J. W. Kim, and C.-C. Kuo, "Quality-of-service mapping mechanism for packet video in differentiated services network," *IEEE Trans. Multimedia*, vol. 3, no. 2, pp. 219–231, June 2001.

[76] D. Quaglia and J. C. De Martin, "Adaptive packet classification for constant perceptual quality of service delivery of video streams over time-varying networks," in *Proc. ICME*, vol. 3, Baltimore, MD, July 2003, pp. 369–372.

[77] A. Sehgal and P. A. Chou, "Cost-distortion optimized streaming media over DiffServ networks," in *Proc. ICME*, vol. 1, Lausanne, Switzerland, Aug. 2002, pp. 857–860.

[78] C. E. Luna, Y. Eisenberg, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Joint source coding and packet marking for video transmission over Diffserv networks," presented at the Tyrrhenian Int. Workshop Digital Communications, Capri, Italy, Sept. 2002.

[79] G. Cheung, W. Tan, and T. Yoshimura, "Rate-distortion optimized application-level retransmission using streaming agent for video streaming over 3G wireless network," in *Proc. ICIP*, Rochester, NY, 2002, pp. I 529–I 532.

[80] J. Chakareski and P. A. Chou, "Application layer error correction coding for rate-distortion optimized streaming to wireless clients," presented at the ICASSP, Orlando, FL, 2002.

[81] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Rate-distortion optimized product code forward error correction for video transmission over IP-based wireless networks," presented at the ICASSP, QC, Canada, May 2004.

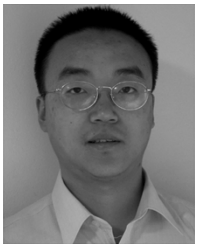
Aggelos K. Katsaggelos (Fellow, IEEE) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1979 and the M.S. and Ph.D. degrees both in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1981 and 1985, respectively.

In 1985, he joined the Department of Electrical and Computer Engineering at Northwestern University, Evanston, IL, where he is currently Professor, holding the Ameritech Chair of Information Technology, and the Director of the Motorola Center for Communications.

Dr. Katsaggelos is a member of the Publication Board of the PROCEEDINGS OF THE IEEE, the IEEE Technical Committees on Visual Signal Processing and Communications, and Multimedia Signal Processing. He has served as editor-in-chief of the *IEEE Signal Processing Magazine* (1997–2002), an Associate editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1990–1992), an area editor for *Graphical Models and Image Processing* (1992–1995), 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 Multi-Dimensional Signal Processing (1992–1998). 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 co-inventor of nine international patents and the recipient of the IEEE Third Millennium Medal (2000), and the IEEE Signal Processing Society Meritorious Service Award (2001).

Yiftach Eisenberg (Member, IEEE) received the B.S. degree in electrical engineering from the University of Illinois at Urbana-Champaign in 1999. He received the M.S. and Ph.D. degrees in electrical engineering from Northwestern University, Evanston, IL, in 2001 and 2004, respectively.

In 2000, he was a Visiting Researcher at Motorola Labs., Schaumburg, IL, in the Multimedia Research Laboratory. His primary research interests include: multimedia processing and communications, video quality analysis and enhancement, and information theory.



Fan Zhai (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Nanjing University, Nanjing, Jiangsu, China, in 1996 and 1998, respectively, and the Ph.D. degree in electrical engineering from Northwestern University, Evanston, IL, in 2004.

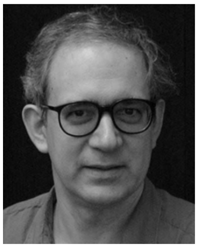
He is currently a Systems Engineer in the Digital Video Department at Texas Instruments, Dallas, TX.

His research interests include image and video signal processing and compression, multimedia communications and networking, and multimedia analysis.

Randall Berry (Member, IEEE) 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, Cambridge, in 1996 and 2000, respectively.

He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at Northwestern University, Evanston, IL. In 1998, he was on the technical staff at Massachusetts Institute of Technology, Lincoln Laboratory in the Advanced Networks Group. His primary research interests include wireless communication, data networks, and information theory.

Dr. Berry is the recipient of a 2003 NSF CAREER award.



Thrasyvoulos N. Pappas (Senior Memebr, IEEE) received the S.B., S.M., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 1979, 1982, and 1987, respectively.

From 1987–1999, he was a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ. In September 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 video compression, video transmission over packet-switched networks, perceptual models for image processing, model-based halftoning, image and video analysis, video processing for sensor networks, audiovisual signal processing, and DNA-based digital signal processing.

Dr. Pappas has served as chair of the IEEE Image and Multidimensional Signal Processing Technical Committee, Associate Editor and Electronic Abstracts Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, technical program co-chair of ICIP-01 and the Symposium on Information Processing in Sensor Networks (IPSN-04), and since 1997 he has been co-chair of the SPIE/IS&T Conference on Human Vision and Electronic Imaging. He is also co-chair for the 2005 IS&T/SPIE Symposium on Electronic Imaging: Science and Technology.