VAPOR: Variance-Aware Per-Pixel Optimal Resource Allocation

Yiftach Eisenberg, *Member, IEEE*, Fan Zhai, *Member, IEEE*, Thrasyvoulos N. Pappas, *Senior Member, IEEE*, Randall Berry, *Member, IEEE*, and Aggelos K. Katsaggelos, *Fellow, IEEE*

***Abstract*—Characterizing the video quality seen by an end-user is a critical component of any video transmission system. In packet-based communication systems, such as wireless channels or the Internet, packet delivery is not guaranteed. Therefore, from the point-of-view of the transmitter, the distortion at the receiver is a random variable. Traditional approaches have primarily focused on minimizing the expected value of the end-to-end distortion. This paper explores the benefits of accounting for not only the mean, but also the variance of the end-to-end distortion when allocating limited source and channel resources. By accounting for the variance of the distortion, the proposed approach increases the reliability of the system by making it more likely that what the end-user sees, closely resembles the mean end-to-end distortion calculated at the transmitter. Experimental results demonstrate that variance-aware resource allocation can help limit error propagation and is more robust to channel-mismatch than approaches whose goal is to strictly minimize the expected distortion.**

***Index Terms*—Error analysis, multimedia communication, teleconferencing, video coding, videophone systems.**

I. Introduction

NETWORK-BASED video applications have dramatically increased in popularity in recent years. One of the major challenges in supporting these applications is that packet delivery is not guaranteed in most practical transmission systems, such as wireless channels and the Internet. Thus, from the point-of-view of the transmitter, the distortion at the receiver is a random variable that depends on the probability of packet loss in the channel. This paper explores the benefits of accounting for the variance, as well as the mean of the end-to-end distortion when allocating limited source and channel resources. By accounting for the variance of the distortion, the proposed approach increases the likelihood that what the end-user sees closely resembles the mean end-to-end distortion calculated at the transmitter, thus increasing the reliability of the system.

Recent work on resilient video coding and transmission for packet lossy networks has primarily focused on improving the end-to-end video quality [1]–[10]. These techniques account for the distortion due to both source compression and channel errors. A common feature among these works is that they all measure video quality by the expected distortion, where the expectation is computed with respect to the probability of packet loss. The work presented here differs from these approaches in that the variance of the distortion, as well as the expected distortion, is used to evaluate video quality.

Several methods have been proposed for calculating the expected distortion. These methods can be divided into two general categories. The first is optimal per-pixel estimation methods, such as [1]–[3]. 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. For example, in [1], Hind's method recursively calculates the distribution of each reconstructed pixel value. In [2], the authors present a recursively accurate method for computing the expected mean absolute difference. In [3], Zhang *et al.* develop an algorithm called ROPE, which efficiently computes the expected mean-squared error (MSE) by recursively calculating only the first and second moment of each pixel. In many advanced video coding schemes, e.g., H.264/AVC 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 [4], [5].

The second category of distortion estimation techniques consists of schemes that use models to estimate the expected distortion, e.g., [6]–[10]. Model-based methods are useful when computational complexity and storage capacity are limited. In [6], the authors present a recursive distortion estimation algorithm, which only differs slightly from ROPE in that they approximate the distortion due to concealment. In order to estimate the expected distortion, a likely subset of the possible loss patterns is considered in [7]. In [8], a limited number of channel simulations and decoder reconstructions are stored at the encoder in order to estimate the expected distortion. In [9], a model is developed for estimating both source and channel distortion based on the intrarefresh rate and the percentage of zeros among the quantized transform coefficients. Another popular metric for calculating the expected distortion is to consider the reduction in distortion given that a packet and all the packets it depends on are received correctly, as in [10]. This approach works well when the dependencies between packets are clearly defined, e.g., when transmitting pre-encoded sequences, but it may not be well suited for real-time applications.

Manuscript received January 16, 2004; revised December 27, 2004. This work was supported in part by the National Science Foundation under Grant CCR-0311838. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Fernando M. B. Pereira.

Y. Eisenberg is with BAE SYSTEMS, Nashua, NH 03061 USA (e-mail: yiftach.eisenberg@baesystems.com).

F. Zhai is with Texas Instruments, Dallas, TX 75243 USA (e-mail: fzhai@ti.com).

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

Digital Object Identifier 10.1109/TIP.2005.860600

Fig. 1. (a) Expected frame and (b), (c) two channel loss simulations.

At the receiver, the end user sees only one of many possible reconstructed sequences, depending on which packets are actually lost. Therefore, the actual distortion at the receiver does not necessarily equal the expected distortion. To illustrate this point, consider the images shown in Fig. 1. While the expected reconstructed frame (averaged over all possible loss realizations) may be reasonable, as shown in Fig. 1(a), the quality at the receiver may vary greatly based on which packets are lost [Fig. 1(b) and (c)]. Therefore, in this paper, we argue that the variance of the end-to-end distortion should also be considered when characterizing video quality in lossy packet networks. We introduce the concept of "Variance-aware per-pixel optimal resource-allocation" (VAPOR) and present a framework for controlling both the expected value and the variance of the end-to-end distortion. In Section VI, experimental results demonstrate that reducing the variance of the distortion can help limit the propagation of perceptually annoying artifacts. In addition, VAPOR is shown to be more robust to channel-mismatch than approaches whose goal is to strictly minimize the expected distortion. This paper builds on our prior work, some of which can be found in [11] and [12].

In addition to the variance in distortion caused by packet loss, there are other sources of quality variation that have received considerable attention in the literature. For example, *rate-control* schemes, i.e., assigning bandwidth (bits) to the different frames in a video sequence, are responsible for the problem of temporal variations in quality [13], [14]. Similarly, approaches such as [15]–[17] have looked at the benefits of providing more even quality across a group of frames. Reducing the spatial variation in quality within a frame has also been considered. The motivation behind these approaches is to prevent having some regions of a frame with excellent quality and others with relatively poor quality. In [18] and [16], one attempt at producing more even quality was to minimize the maximum distortion within a frame as opposed to the average distortion. The approaches mentioned above are complementary to the work presented here. Our approach addresses another source of quality variation, i.e., the variance in distortion caused by random channel losses.

The concept of variance-aware resource allocation is pertinent to a wide range of video transmission systems including wireless communications, wired networks, and Differentiated Services (DiffServ) Internet systems. In our experimental results, we consider wireless video communications as a representative example. The basic ideas presented here can be utilized in various applications such as video conferencing, target tracking, surveillance, and personal communications to name a few.

The remainder of the paper is organized as follows. Next, we describe the system model. In Section III, we characterize the end-to-end distortion in packet-based video communication systems. Two variance-aware resource allocation formulations are presented in Section IV, followed by a discussion of the solution approach in Section V. Extensive experimental analysis is presented in Section VI. Section VII contains concluding remarks.

II. System Model

Consider a packet-based video communication system where the video is encoded using a block-based motion-compensated technique (e.g., H.263 [19], MPEG-4 [20]). Each frame is divided into slices that are comprised of consecutive macro-blocks (MBs). Each slice is independently decodable, i.e., the decoding of one slice is not affected by the loss of other slices in the same frame. Losses in other frames, on the other hand, may cause temporal error propagation due to interframe prediction. After a slice is encoded, it is transmitted across a 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 packet, source-coding parameters, such as the coding mode (intra/inter/skip) and quantization step-size for each MB are specified. We use $s^{n,k}$ to denote the source-coding parameters for the kth packet in the nth frame, and $s^n = \{s^{n,1}, \ldots, s^{n,K}\}$ to denote the coding parameters for all the packets in the frame. The number of bits used to encode the kth packet $B^{n,k}$ is a function of $s^{n,k}$; we use $B^{n,k}(s^{n,k})$ to explicitly indicate this dependency.

Error resilient video coding is a well established field of research and includes a broad range of techniques, such as adaptive mode and quantizer selection, data partitioning, reversible variable length coding, and various packetization schemes. In this paper, our focus is on resilient coding algorithms which adapt the prediction mode (and quantization parameters) for each MB in the sequence. In other words, given the bit budget constraints imposed by the application, these techniques address the problem of trading off the compression efficiency of intercoding with the resilience to error propagation of intracoding. For more details on other resilient source coding techniques we refer the reader to [24].

After each packet is encoded, it is transmitted over a communication channel. Most practical communication systems, such as wired and wireless networks, are unreliable, and, therefore, packets are lost during transmission with some probability.[1] An implicit assumption in our system model is that the probability of loss in the channel can be reasonably estimated at the transmitter and is, therefore, known when allocating limited source and channel resources. Channel estimation is a well established and sophisticated field of research. Many techniques and protocols have been developed to estimate the probability of loss including the use of direct feedback from the receiver or intermediate nodes in the network, measuring the signal strength of a pilot signal in a wireless channel, or through analytical models. In Section VI-A, we provide a model for characterizing the probability of loss in a wireless channel based on the concept of "outage capacity" developed in [26]. In Section VI-D, we discuss the robustness of the proposed approach to channel mismatch, i.e., when the estimate of the channel loss rate is incorrect.

[1]Note that, here, we assume that packets are either received error free or considered lost due to packet corruption or excessive delay.

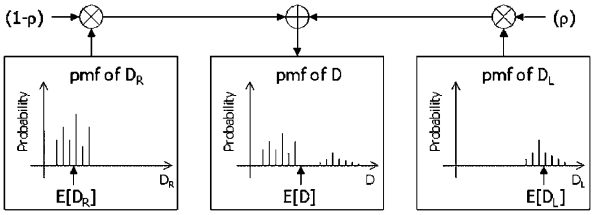

Fig. 2. Example of the probability mass functions of $D^{n,k,i}$, $D_R^{n,k,i}$, $D_L^{n,k,i}$, and their relationship.

III. Distortion Characterization

In this section, we present a framework for characterizing the distortion between the original video and the reconstructed sequence at the receiver. This framework is based on knowledge of how the original sequence is encoded, the probability of packet loss in the channel, and the concealment strategy used by the decoder. Since a pixel is the smallest information symbol in a digital video sequence, we use per-pixel accurate calculations to characterize the end-to-end distortion.

Consider a single pixel in the video sequence. For the system described in Section II, we assume the encoded information for each pixel is contained in only one packet and that each packet is either received correctly or lost. This means for example, we assume no layered coding although the concept of variance aware resource allocation can be extended to such a scenario. In this case, the distortion between the original and reconstructed value for the ith pixel of the kth packet in the nth frame is a random variable with the following distribution

$$D^{n,k,i} = \begin{cases} D_R^{n,k,i}, & \text{with probability } (1-\rho^{n,k}) \\ D_L^{n,k,i}, & \text{with probability } (\rho^{n,k}) \end{cases} \quad (1)$$

where $\rho^{n,k}$ is the probability that the kth packet is lost, $D_R^{n,k,i}$ is the distortion if the packet is received correctly, and $D_L^{n,k,i}$ is the distortion if the packet is lost. If the pixel is predictively encoded, then $D_R^{n,k,i}$ is a random variable due to random losses in previous frames. On the other hand, if the pixel is independently encoded (intracoded), then $D_R^{n,k,i}$ is deterministic. The distortion if a packet is lost depends on the concealment strategy used at the decoder. If prediction is used in the concealment strategy, e.g., temporal concealment, then $D_L^{n,k,i}$ is also a random variable. The probability mass function (pmf) of the end-to-end distortion for a given pixel $D^{n,k,i}$ and its relationship to the quantities discussed above are shown in Fig. 2. As shown, the pmf of $D^{n,k,i}$ is simply the weighted sum of the pmfs of $D_R^{n,k,i}$ and $D_L^{n,k,i}$, where the weight is controlled by the probability of loss $\rho^{n,k}$.

A. Expected Value of the End-to-End Distortion

The expected value of the end-to-end distortion for a given pixel can be written as

$$E\left[D^{n,k,i}\right] = (1-\rho^{n,k})E\left[D_R^{n,k,i}\right] + (\rho^{n,k})E\left[D_L^{n,k,i}\right] \quad (2)$$

where $E[\bullet]$ indicates the expected value taken with respect to the probability of loss. The expected distortion for the kth packet in the nth frame is defined as

$$\overline{D}^{n,k} = \frac{1}{I^k}\sum_{i=1}^{I^k} E[D^{n,k,i}] \quad (3)$$

where I^k is the number of pixels in the kth packet. Similarly, the average expected distortion for the nth frame and for the sequence are simply

$$\overline{D}^{n} = \frac{1}{M}\sum_{k=1}^{K} (I^k)\overline{D}^{n,k} \quad (4)$$

and

$$\overline{D}_{\text{seq}} = \frac{1}{N}\sum_{n=1}^{N} \overline{D}^{n} \quad (5)$$

respectively, where K is the number of packets in a given frame, M is the number of pixels per frame, and N is the number of frames in the sequence.

One way to minimize the expected distortion for a given frame $\overline{D}^n$ is by efficiently allocating source coding resources, e.g., source bits. For a given bit budget, intercoding typically has higher compression efficiency than intracoding, but is more susceptible to temporal error propagation. The expected distortion accounts for the distortion due to both source coding and error propagation. Therefore, if the objective is to minimize the expected distortion, and the increase in source distortion caused by intracoding outweighs the increase in error propagation distortion caused by intercoding, then it is better to use intercoding for the given MB. In Section VI, we show that under certain conditions, e.g., low to moderate probability of packet loss, the prevailing selection of intercoding over intra may lead to prolonged propagation of perceptually visible artifacts.

B. Variance of the End-to-End Distortion

In a video transmission application, the actual distortion in the reconstructed sequence (i.e., for a single loss realization) is not equal to the expected distortion. Therefore, to determine the likelihood that the actual distortion seen by the end-user is near the expected distortion, one must look at the variance of the distortion.

The variance of the distortion for a given pixel is by definition equal to $\text{Var}[D^{n,k,i}] = E[(D^{n,k,i})^2] - E[D^{n,k,i}]^2$. By substituting (1) into the previous equation and rearranging terms, we can express $\text{Var}[D^{n,k,i}]$ as

$$\begin{aligned}\text{Var}[D^{n,k,i}] = {} & (1-\rho^{n,k})\text{Var}\left[D_R^{n,k,i}\right] \\ & + (\rho^{n,k})\text{Var}\left[D_L^{n,k,i}\right] + (1-\rho^{n,k})(\rho^{n,k}) \\ & \times \left\{E\left[D_R^{n,k,i}\right] - E\left[D_L^{n,k,i}\right]\right\}^2 \end{aligned} \quad (6)$$

where $\text{Var}\left[D_R^{n,k,i}\right]$ and $\text{Var}\left[D_L^{n,k,i}\right]$ are the variance in distortion given that the packet containing the coding information for this pixel is received and lost, respectively.

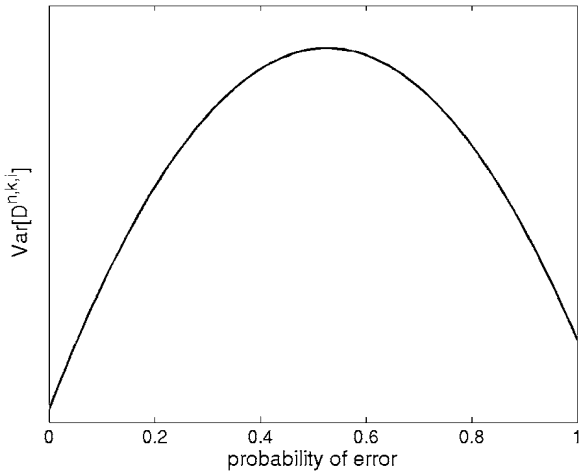

Fig. 3. Relationship between the variance in distortion for a given pixel $\mathrm{Var}\left[D^{n,k,i}\right]$ and the probability of packet loss $\rho^{n,k}$.

As expected, $\mathrm{Var}[D^{n,k,i}]$ increases when $\mathrm{Var}\left[D_R^{n,k,i}\right]$ or $\mathrm{Var}\left[D_L^{n,k,i}\right]$ increase. Therefore, intracoding, which has $\mathrm{Var}\left[D_R^{n,k,i}\right] = 0$, enables the transmitter to greatly decrease the variance. Intercoding on the other hand is susceptible to temporal error propagation and, thus, has $\mathrm{Var}\left[D_R^{n,k,i}\right] \geq 0$. This suggests that $\mathrm{Var}\left[D_R^{n,k,i}\right]$ is a good indicator of how severely a packet may be affected by error propagation. This important observation is discussed further in Section VI-C. From (6), we see that $\mathrm{Var}\left[D^{n,k,i}\right]$ increases as $\left\{E\left[D_R^{n,k,i}\right] - E\left[D_L^{n,k,i}\right]\right\}^2$ increases. This means that if a pixel is difficult to conceal, i.e., $E\left[D_L^{n,k,i}\right] \gg E\left[D_R^{n,k,i}\right]$, its distortion varies greatly depending on whether the packet is received or lost.

From (6), it can be seen that $\mathrm{Var}[D^{n,k,i}]$ is a concave quadratic function of $\rho^{n,k}$, with a maximum at

$$\rho^{n,k} = \frac{1}{2} + \frac{\left\{\mathrm{Var}\left[D_L^{n,k,i}\right] - \mathrm{Var}\left[D_R^{n,k,i}\right]\right\}}{2\left\{E\left[D_R^{n,k,i}\right] - E\left[D_L^{n,k,i}\right]\right\}^2} \tag{7}$$

as shown in Fig. 3. This is intuitively satisfying since there is less variability in $D^{n,k,i}$ when the probability of loss is either very small or very large. For example, if $\rho = 0$ or 1 for all the packets in the sequence, then from the point of view of the transmitter the reconstructed sequence at the decoder is no longer random. In this extreme case, $\mathrm{Var}[D^{n,k,i}] = \mathrm{Var}\left[D_R^{n,k,i}\right] = \mathrm{Var}\left[D_L^{n,k,i}\right] = 0$.

Let $\mathrm{Std}[D^{n,k,i}] = \sqrt{\mathrm{Var}[D^{n,k,i}]}$ represent the standard deviation of the distortion for a given pixel. The average standard deviation in distortion for the kth packet in the nth frame is defined as

$$\overline{\sigma}^{n,k} = \frac{1}{I^k}\sum_{i=1}^{I^k} \mathrm{Std}[D^{n,k,i}]. \tag{8}$$

Similarly, the average standard deviation in distortion for the nth frame and for the sequence are respectively

$$\overline{\sigma}^{n} = \frac{1}{M}\sum_{k=1}^{K} (I^k)\overline{\sigma}^{n,k} \tag{9}$$

and

$$\overline{\sigma}_{\mathrm{seq}} = \frac{1}{N}\sum_{n=1}^{N} \overline{\sigma}^{n}. \tag{10}$$

It is important to note that the average variance in distortion per pixel as in (9) is not equal to the variance of the average distortion for a frame, i.e.,

$$\frac{1}{M}\sum_{k=1}^{K}\sum_{i=1}^{I^k} \mathrm{Var}[D^{n,k,i}] \neq \mathrm{Var}\left[\frac{1}{M}\sum_{k=1}^{K}\sum_{i=1}^{I^k} D^{n,k,i}\right]. \tag{11}$$

The expression on the right hand side of (11) is more difficult to compute since it requires calculating the cross-correlation between the distortion of all the pixels in the frame. In addition, the problem of optimally encoding and transmitting a video frame based on the variance in quality, as done in Section IV, becomes infeasible if the $\mathrm{Var}\left[(1/M)\sum_{k=1}^{K}\sum_{i=1}^{I^k} D^{n,k,i}\right]$ is used. Another motivation for using (9) is that it captures local variations in quality better than the variance of the average distortion.

C. Squared Error Distortion Metric

The expressions for the mean (2) and variance (6) of the end-to-end distortion derived in Sections I and II hold for a wide range of distortion metrics. In our experimental results, we consider the case where distortion is defined as the squared error between the original pixel value x and the reconstructed pixel value at the receiver $\widetilde{x}$. In this case, the first two moments of the reconstructed pixel value, i.e., $E[\widetilde{x}]$ and $E[(\widetilde{x})^2]$, are needed to accurately calculate the expected distortion $E[D] = E[(x - \widetilde{x})^2]$. Similarly, the first four moments of the reconstructed pixel value are needed to accurately calculate $\mathrm{Var}[D] = E[(x - \widetilde{x})^4] - E[(x - \widetilde{x})^2]^2$.

In certain cases, optimal distortion estimation methods, such as ROPE [3] and [1] can be used to accurately and recursively calculate the necessary reconstructed pixel moments. In other cases, such as noninteger pixel motion compensation, models may be needed to estimate the mean and variance of the distortion. Developing efficient models for estimating the variance of the end-to-end distortion is an area requiring future research.

IV. Variance-Aware Resource Allocation

Our objective is to efficiently allocate limited resources at the transmitter in order to minimize the end-to-end distortion while satisfying the delay constraints imposed by the application. The specific parameters which can be adapted at the transmitter depend on the application. For example, in wireless communications, the transmitter may be able to adapt how the source is encoded, as well as the transmission power and modulation scheme. Examples of other control parameters include the channel coding rate or the packet classification in a DiffServ Internet system. To better illustrate the potential benefits of the

proposed approach, we consider the problem of optimally controlling the source coding parameters $S^{n,k}$, i.e., the prediction mode and quantizer. It should be noted though that the concepts developed here can be used in a wide range of applications in which both source and channel resources may be adapted.

We can now formally state our objective as

$$\min_{\{s^{n,k}\}} \quad D^n_{\text{tot}}(s^n) \tag{12.1}$$

$$\text{subject to}: \quad T^n_{\text{tot}} = \sum_{k=1}^{K} \frac{B^{n,k}(s^{n,k})}{R} \leq T^n_0 \tag{12.2}$$

where D^n_{tot} is the cost function representing the end-to-end distortion, T^n_{tot} is the total transmission delay, and T^n_0 is the maximum transmission delay for the nth frame. In this paper, we assume that the transmission rate R is fixed, although the formulation can easily be extended to allow variable transmission rate per packet.

We consider a real-time application in which there is a limited amount of time between when a frame is captured at the transmitter and when it must be displayed at the receiver. To manage the delay requirements imposed by the application, we assume that a higher-level controller assigns a transmission delay constraint T^n_0 to each frame in the sequence. The design of this controller is outside the scope of this paper. The value of T^n_0 may vary from frame to frame, but it is assumed to be a known constant when optimizing each frame. Note that when the transmission rate is fixed, the delay constraint in (12.2) can also be viewed as an equivalent bit budget constraint.

In previous sections, we discussed the need to account for both the mean and the variance of the end-to-end distortion when evaluating video quality. One way to do this is by defining the cost function to be minimized in (12.1) as the weighted sum of the expected distortion plus the standard deviation in distortion, i.e.,

$$\begin{aligned} D^n_{\text{tot}} &= (1-\alpha)\overline{D}^n + (\alpha)\overline{\sigma}^n \\ &= \frac{1}{M}\sum_{k=1}^{K}\sum_{i=1}^{I^k}(1-\alpha)E[D^{n,k,i}] + (\alpha)\text{Std}[D^{n,k,i}] \end{aligned} \tag{13}$$

where $\alpha \in [0,1]$ defines the relative importance of the variance of the end-to-end distortion. We use $\text{Std}[D^{n,k,i}]$ instead of $\text{Var}[D^{n,k,i}]$ so that the units of D^n_{tot} are consistent. By using (13) as the objective in the optimization problem (12) and solving for different values of α, we can observe the tradeoff between minimizing the expected value and the variance of the distortion. For example, in Fig. 4, we plot the mean and the standard deviation in distortion versus α for frame 43 of the "Foreman" test sequence. As expected, increasing α reduces the standard deviation at the cost of increased expected distortion. When $\alpha = 0$, we obtain a special case of the general formulation in which the objective is to strictly minimize the expected distortion per frame, as in [1]–[10]. In Section VI, this special case is referred to as the minimum expected distortion (MED) approach. As shown in Fig. 4, it may be possible to significantly reduce $\overline{\sigma}^n$ while only slightly increasing $\overline{D}^n$.[2] This observation motivates the following alternative variance-aware formulation.

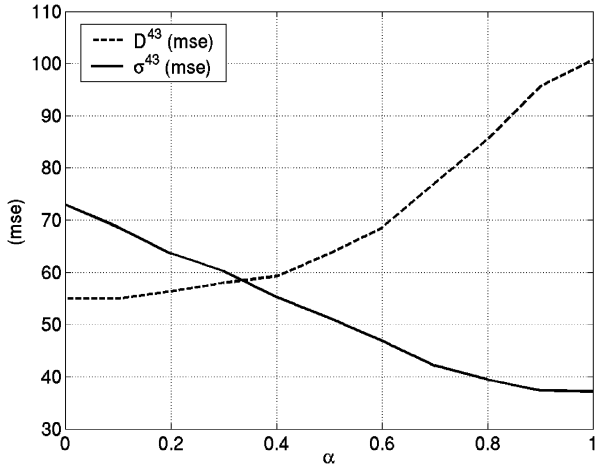

Fig. 4. Impact of α on the average expected distortion ($\overline{D}^{43}$) and average standard deviation in distortion ($\overline{\sigma}^{43}$) for frame 43 of the "Foreman" sequence coded at 30 fps with $R = 150$ kbps and $\rho = 0.01$.

By adjusting α, we can control the value of both the mean and the variance of the distortion. Let $\overline{D}^n$ be the MED for the frame (i.e., the solution to (12) with $\overline{D}^n_{\min}$). We modify (12) slightly to include the optimal selection of α, by constraining the difference between the expected distortion and the minimum achievable value . In other words, we solve the following optimization:

$$\min_{\{s^{n,k},\alpha\}} \quad D_{\text{tot}} = (1-\alpha)\overline{D}^n + (\alpha)\overline{\sigma}^n \tag{14.1}$$

$$\text{subject to}: \quad T^n_{\text{tot}} = \sum_{k=1}^{K} \frac{B^{n,k}(s^{n,k})}{R} \leq T^n_0 \tag{14.2}$$

and

$$\overline{D}^n \leq \Delta_d \times \overline{D}^n_{\min} \tag{14.3}$$

where $\Delta_d \geq 1$ represents the maximum increase in expected distortion we are willing to tolerate in order to decrease the variance of the distortion. Note that the formulation in (14) is equivalent to minimizing the variance given an expected distortion constraint. In Section VI, we show that for a small increase in expected distortion, the standard deviation of the distortion can be reduced significantly.

To the best of our knowledge, the formulations in (12) and (14) are the first to account for both the mean and the variance of the end-to-end distortion. We refer to formulations of this type as a "variance-aware per-pixel optimal resource-allocation" (VAPOR) techniques. The formulation in (12) is less complex than (14) in that α is fixed. The drawback is that it may not be very intuitive how to set α, and the same α may not be desirable for every frame. This problem is addressed in (14) because α is optimally set based on a specified tolerable increase in expected distortion.

[2]Similar results have been obtained using other sequences. Note that the extent to which the standard deviation decreases relative to the expected distortion depends on the signal content as well as the parameter settings.

V. Solution Approach

In order to solve (12), we use Lagrange relaxation and dynamic programming (DP). First, we introduce a Lagrange multiplier $\lambda \geq 0$ and solve the relaxed problem

$$\min_{\{s^{n,k}\}} D^n_{\text{tot}}(s^n) + \lambda T^n_{\text{tot}}(s^n). \tag{15}$$

By appropriately choosing λ, the solution to (12) can be obtained within a convex-hull approximation by solving (15) [21], [22]. Various methods, such as cutting-plane or sub-gradient methods, can be used to search for λ [23]. In our experimental results, we use a simple yet effective bisection method.

For each choice of λ, we can solve (15) using DP. Predictive coding and error concealment both introduce dependencies between packets. For example, temporal concealment based on the motion vectors (MVs) of neighboring packets causes the distortion for a given packet to depend on how neighboring packet(s) in the same frame are encoded as well as their probability of loss. Accounting for the dependencies between packets is what makes solving (15) a difficult problem. DP can be used to efficiently find the optimal $s^{n,k}$ for each packet in the frame when the dependencies between packets are limited, e.g., to a small neighborhood. For more details, please see [11], [21], and [22]. To solve (14), we can iteratively adjust α and solve (12) until the resulting expected distortion $\overline{D}^n$ is less than or equal to $\Delta_d \times \overline{D}^n_{\min}$.

VI. Experimental Results

The basic ideas developed throughout this paper are relevant to a wide range of applications. In this section, we focus on one example, i.e., real-time wireless video communications. Experimental results highlight the potential benefits of accounting for both the expected value and the variance of the distortion when allocating limited resources in packet-based video transmission systems. As a comparison to the proposed VAPOR approach, we consider a more traditional approach whose goal is to minimize the expected distortion per frame, as in [1]–[10]. We refer to this scheme as the MED approach. As mentioned in Section IV, the MED approach is a special case of (12) in which $\alpha = 0$.

In Section VI-B, we consider the problem of optimal source coding assuming that the probability of loss in the channel is known or can be reasonably estimated at the transmitter. In Sections VI-C and VI-D, we present results suggesting that VAPOR can help reduce error propagation and channel mismatch sensitivity.

A. Experimental Setup

Source Coding: An H.263+ [19] codec is used to encode QCIF resolution video sequences using a limited number of quantization step sizes for "intra" and "inter" MBs. In addition, integer pixel motion compensation is used in order to ensure accurate distortion estimation at the transmitter. A thorough discussion on the necessary conditions for accurately calculating the expected distortion at the transmitter using methods such as ROPE can be found in [3]. As in [11], we consider the case where each packet contains one MB, i.e., each MB is independently decodable. Although this packetization scheme has low coding efficiency, it is highly resilient to channel errors [24], [25].

Similarly, we consider a relatively simple concealment strategy in which the concealment MV for a lost MB is defined as the MV of the MB to the left of the lost MB. If the lost MB is on the left edge of the frame, or if the MB to the left is also lost, then the concealment MV is set to zero.

We consider a real-time application with an allowable transmission delay of one frame duration. Therefore, a sequence coded at 30 frames per second (fps) has a delay constraint $T_0 = 33$ ms. At 15 fps, $T_0 = 66$ ms. Since the transmission rate is fixed, the delay constraint imposed by the application translates to an equivalent bit budget constraint per frame. In other words, all the approaches considered here use roughly the same number of bits when encoding a given frame in the sequence. The only difference between the various approaches is how they encode the frames, i.e., what mode and quantizer they use to encode each MB. In all the experiments, the "generalized skip mode," introduced in [11], enables the transmitter to intentionally not transmit certain packets if their concealment at the decoder results in adequate quality. Note that the decision not to transmit a packet is made within the optimization framework and can be viewed as one of the possible coding modes for each packet.

The proposed VAPOR approach is applicable to a wide range of coding and transmission schemes including wireless video communications and Interne-based video transmission. Therefore, it is important to note that the experimental setup is chosen to illustrate the concepts introduced in this paper and can easily be adapted based on the application and system requirements. For example, the formulations in (12) and (14) can be used with other packetization schemes in which several MBs or even entire frames are placed in each packet. Selecting the optimal packetization scheme is application specific and depends on many factors including the source content, the available channel bandwidth, and how source packets are repacketized at other layers in the network protocol stack. Smaller packets provide more flexibility and resilience to errors but have more overhead. Larger packets provide improved coding efficiency, but cause large localized distortions when they are lost. Balancing the tradeoff between coding efficiency and error resilience by optimally selecting the packetization scheme is an important topic, but is outside the scope of this paper. Here our goal is to propose a new variance-aware approach to video communications. This framework can be used to design and evaluate the performance of future video transmission systems.

Channel Model: We use the wireless channel model from [11], where each packet is sent over a flat (frequency nonselective) slowly fading channel with additive white Gaussian noise. We model $\rho^{n,k}$ in the capacity versus outage framework introduced in [26]. For this channel model

$$\rho^{n,k} = 1 - \exp\left(\frac{N_o W}{P^{n,k} E[H^{n,k}]}(2^{R/W} - 1)\right) \tag{16}$$

where W is the bandwidth, $N_o W$ is the noise power, R is the transmission rate in source bits per second, $P^{n,k}$ is the transmission power for the kth packet, and $E[H^{n,k}]$ is the expected value of the channel fading level for the kth packet $H^{n,k}$. We assume that the fading is independent identically distributed (i.i.d.) per

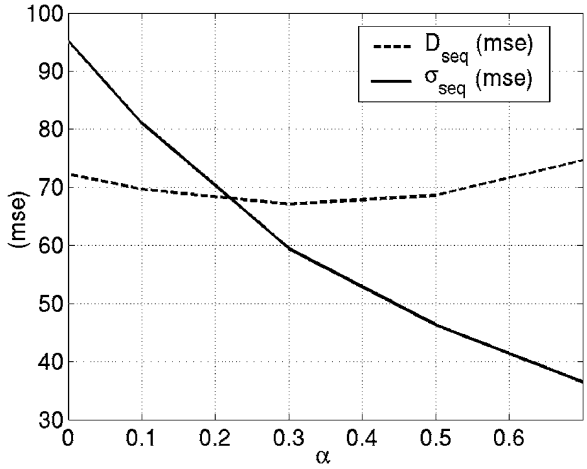

Fig. 5. Impact of α on the expected distortion ($\overline{D}_{\text{seq}}$) and standard deviation in distortion ($\overline{\sigma}_{\text{seq}}$) averaged over all the frames of the "foreman" sequence coded at 30 fps with $R = 150$ kbps and $\rho = 0.01$.

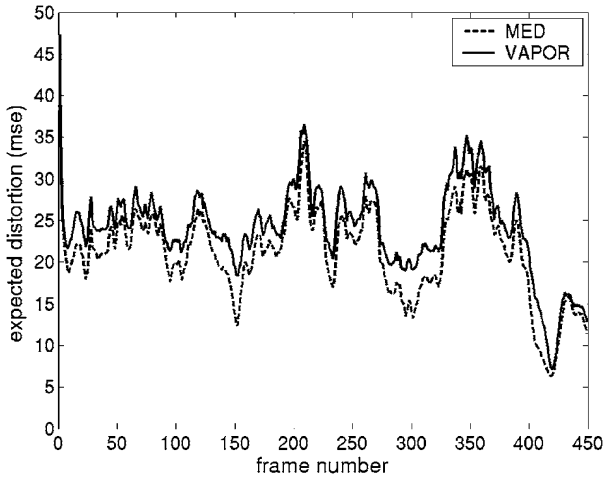

Fig. 6. Expected distortion per frame $\overline{D}^n$ for the "silent" sequence.

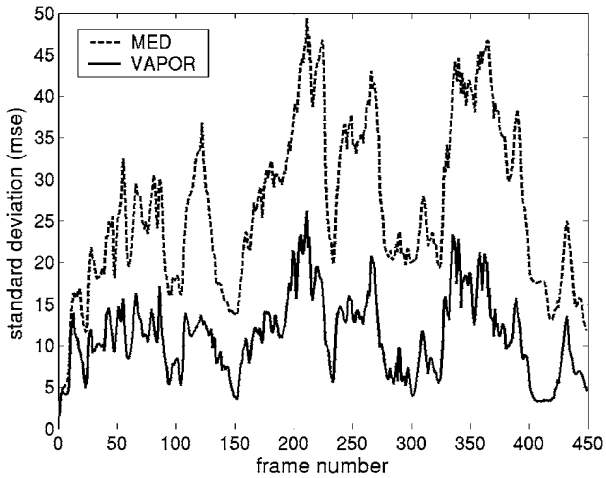

Fig. 7. Average standard deviation in distortion per frame $\overline{\sigma}_n$ for the "silent" sequence.

packet, i.e., $E[H^{n,k}] = E[H]$ for all k.[3] In our experiments, $N_oW/E[H] = 6$ Watts and $W = 5$ MHz. We consider transmission rates ranging from $R = 150$ kbps to 300 kbps. These values are similar to the ones being proposed for next generation wireless standards [27].

In the experiments presented here, we assume that techniques such as packet interleaving can be used to spread out burst error, and thus, model the fading as i.i.d. for the packets in a given frame. A detailed analysis of the robustness of the proposed approaches under bursty error conditions is presented in [28]. The results in [28] show that the performance advantages of the proposed scheme still hold when more complex channel models are used.

B. Error Resilient Source Coding

In this section, we consider the problem of optimal source coding. That is, we focus on how adapting the source coding parameters, such as the prediction mode and quantizer, impacts the end-to-end distortion.

Consider the "Foreman" test sequence coded at 30 fps and transmitted over a 150 kbps channel with probability of packet loss $\rho^{n,k} = 0.01$ for all n, k. In Fig. 5, we plot the expected distortion $\overline{D}_{\text{seq}}(5)$ and the average standard deviation $\overline{\sigma}_{\text{seq}}(10)$ for the sequence as a function of α. In other words, for each fixed value of α we solve (12) and plot the corresponding $\overline{D}_{\text{seq}}$ and $\overline{\sigma}_{\text{seq}}$. The motivation behind this experiment is to study how α affects the statistical properties of the end-to-end distortion for the sequence. As shown in Fig. 5, $\overline{\sigma}_{\text{seq}}$ decreases as α increases. This is intuitively satisfying since a larger α means that more weight is placed on reducing the variance in distortion when allocation resources in (12).

Surprisingly, though, the expected distortion for the sequence $\overline{D}_{\text{seq}}$ does not necessarily increase as α increases, as shown in Fig. 5. Recall that the optimization is carried out on a per frame basis and, therefore, ignores the effects of the current optimization on future frames. Therefore, although setting $\alpha = 0$ results in the lowest $\overline{D}^n$ for a single frame (as shown in Fig. 4), it is not guaranteed that $\overline{D}_{\text{seq}}$ is minimized by myopically minimizing the expected distortion per frame. This result is due to interframe dependencies. Reducing the variance in the current frame may lead to a more reliable prediction for the next frame, which in turn may reduce the overall expected distortion for the sequence.

Recently, there has been work on prescient video coding, i.e., accounting for the effects on future frames when encoding the current frame [29]. The drawback of this work is that a certain number of frames must be captured before the optimization can begin, thus increasing delay. In addition, the reduction in overall distortion gained from optimizing over a group of frames is reported to be relatively small in [29]. Our results suggest that by reducing the variance of the distortion for the current frame we can achieve similar reductions in $\overline{D}_{\text{seq}}$ without the added delay or optimization complexity.

Next, we consider the formulation in (14) where α is adapted per frame based on a tolerable increase in expected distortion Δ_d. Here, we set $\Delta_d = 1.05$ (i.e., a 5% increase). In this experiment, we use the "silent" test sequence coded at 30 fps with $R = 150$ kbps and $\rho^{n,k} = 0.01$. For these settings, the average value of α is 0.64. In Figs. 6 and 7, we compare the

[3]Note that the fading level is assumed to be i.i.d. for all the packets in a single frame, but that the average fading level may vary from frame to frame.

TABLE I
EXPECTED DISTORTION $\overline{D}_{\text{seq}}$ AND STANDARD DEVIATION IN DISTORTION $\overline{\sigma}_{\text{seq}}$ FOR SEVERAL SEQUENCES IN UNITS OF MSE. ALL THE SEQUENCES ARE CODED AT 30 FPS WITH $R = 150$ kbps AND $\rho = 0.01$

	Foreman		Silent		Traffic		Mother Daughter		Susie	
	$\overline{D}_{seq}$	$\overline{\sigma}_{seq}$	$\overline{D}_{seq}$	$\overline{\sigma}_{seq}$	$\overline{D}_{seq}$	$\overline{\sigma}_{seq}$	$\overline{D}_{seq}$	$\overline{\sigma}_{seq}$	$\overline{D}_{seq}$	$\overline{\sigma}_{seq}$
MED	72.19	95.2	21.17	26.6	32.45	66.45	11.58	9.52	21.50	20.10
VAPOR	67.88	48	24.1	11.1	32.41	39.05	12.26	4.69	21.00	12.00

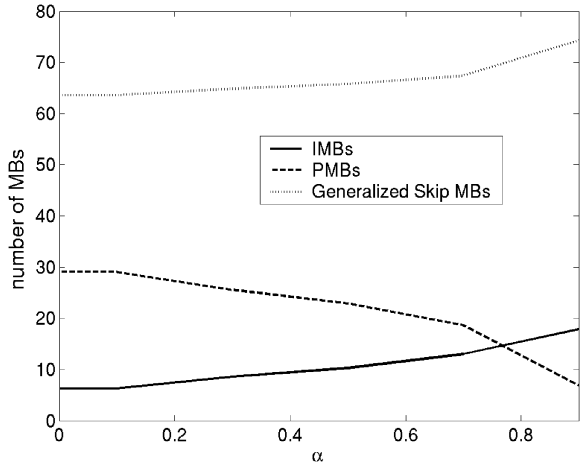

Fig. 8. Mode selection as a function of α for the "silent" sequence with $R = 150$ kbps and $\rho = 0.01$.

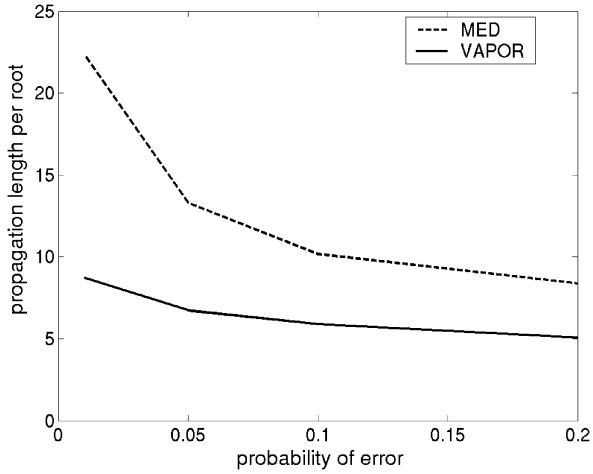

Fig. 9. Average error propagation length per root versus probability of error for the "silent" sequence.

proposed adaptive α approach with the MED approach (i.e., $\alpha = 0$). As shown in Figs. 6 and 7, the expected distortion per frame $\overline{D}^n$(4) is similar for both approaches, while the standard deviation per frame $\overline{\sigma}^n$(9) is significantly smaller for the VAPOR approach. This suggests that by accepting even a small increase in expected distortion it is possible to greatly reduce the variation in quality between different loss realizations. Similar results have been obtained using other sequences. In Table I, we compare the average expected distortion and average standard deviation for several sequences coded using the MED and VAPOR approaches. As seen, the VAPOR approach greatly reduced the standard deviation in distortion while only slightly increasing the expected distortion. As discussed above, for some sequences, such as "Foreman" and "Susie," the expected distortion for the sequence may actually decrease slightly when using VAPOR.

The value of α affects how much of the end-to-end distortion is due to source coding and how much is caused by channel errors. This is primarily a function of mode selection. For each value of α in (12), Fig. 8 shows the average number of MBs per frame that are coded as intra (IMBs), inter (PMBs), or are intentionally not transmitted (generalized skip). As α increases, the number of IMBs increases, as shown in Fig. 8. Since the bit budget is constrained and because IMBs have lower coding efficiency than PMBs, the source coding distortion increases as α increases. On the other hand, increasing the number of IMBs reduces the distortion caused by error propagation. This tradeoff is discussed next.

C. Error Propagation

As in Section VI-B, we focus here on resilient source coding, i.e., mode and quantizer selection given a fixed probability of packet loss in the channel. We consider the "silent" sequence encoded at 30 fps and sent over a 150 kbps channel. The objective in this section is to analyze how sensitive the MED and VAPOR approaches are to error propagation.

Let us define a pixel to be in error if its reconstructed value at the decoder differs from that at the encoder. At the receiver, we can track the temporal propagation of each error in order to identify error propagation paths. Each propagation path has a root, i.e., the origin of the error path, and a length, i.e., the number of pixels in error due to the initial root error. Note that only pixels that are lost and which start a new error path are defined to be root errors. In other words, a lost pixel which propagates a previous error is not considered to be a root error.

The average number of root errors per frame is a function of the probability of packet loss ρ. Since the probability of loss is fixed, both the MED and VAPOR approaches have roughly the same number of root errors per frame. As expected, the average number of roots per frame increases as ρ increases, for both approaches.

In Fig. 9, we plot the average length of error propagation per root as a function of ρ. The results are obtained by averaging over 50 channel loss simulations. As shown, the average number of pixel errors caused by an initial root error is significantly smaller for the VAPOR approach than the MED approach. This is especially true at lower probabilities of loss.

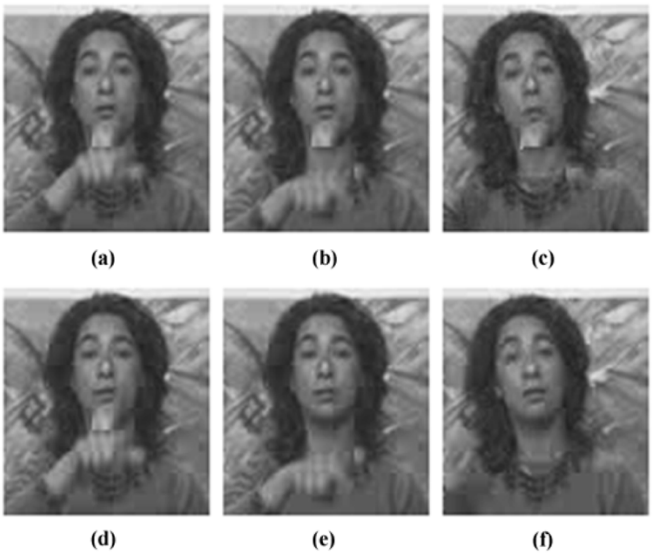

Fig. 10. Error propagation example. MED approach: frame number (a) 109, (b) 110, and (c) 123. VAPOR approach: frame number (d) 109, (e) 110, and (f) 123.

When the probability of loss is low, the expression for the expected distortion (2) is dominated by the expected distortion if the packet is received $E\left[D_R^{n,k,i}\right]$. In addition, at low loss probabilities the reference frame at the decoder is more likely to be correctly reconstructed, and, hence, source coding distortion becomes the primary component of $E\left[D_R^{n,k,i}\right]$. Therefore, as channel conditions improve, an approach whose goal is to minimize the expected distortion will use more intercoding in order to reduce the distortion due to compression (as shown in Fig. 8). The side effect of increased intercoding is susceptibility to prolonged error propagation. This is why the average length of error propagation drastically increases for the MED approach as the probability of loss decreases, as seen in Fig. 9.

At low probability of loss, the expression for the variance in distortion (6) is dominated by $\mathrm{Var}\left[D_R^{n,k,i}\right]$. As discussed in Section III-B, intracoding results in $\mathrm{Var}\left[D_R^{n,k,i}\right] = 0$, while intercoding has $\mathrm{Var}\left[D_R^{n,k,i}\right] \geq 0$. Therefore, a variance-aware resource allocation scheme, such as (12) or (14), uses more intracoding than a MED approach in order to reduce the variance in quality if a packet is received, as shown in Fig. 8. Perceptually, the increased number of intra MBs results in faster termination of error propagation.

In Fig. 10, we compare a series of reconstructed frames at the decoder for the MED and VAPOR approaches. Note that these images are for a single channel loss simulation, i.e., the same MBs are lost in both schemes. As shown, both approaches suffer a loss in frame 109 where the woman's hand goes across her chin. The difference between the two approaches is that in frame 110, the VAPOR approach intra refreshes this region while the MED approach does not. Thus, this error persists till frame 123 in the MED approach while it has been quickly removed by VAPOR. It is important to note that no feedback is used in either approach and that the difference in mode selection is purely due to estimates of the mean and variance of the end-to-end distortion.

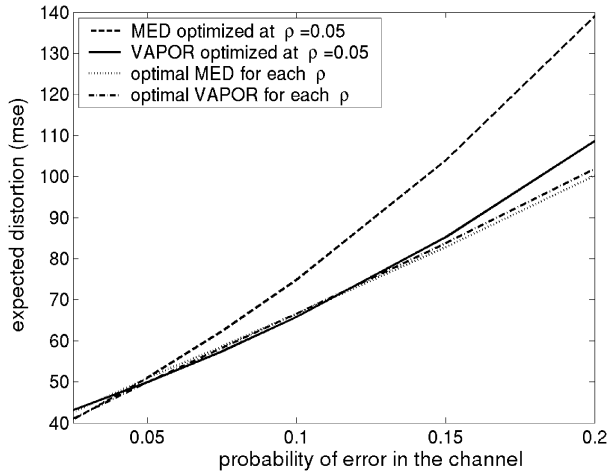

Fig. 11. Channel mismatch resilience for the "Foreman" sequence.

From a communications point of view, transmitting information at the lowest possible probability of error is desirable. The point of the above discussion is that at this desired operating point, an approach whose goal is to minimize the expected distortion may become susceptible to prolonged error propagation. A variance-aware approach on the other hand helps prevent prolonged error propagation because it accounts for the variability in quality caused by error propagation. As shown in Fig. 9, VAPOR reduces error propagation even at higher probabilities of packet loss. Thus, we argue that to be more resilient to error propagation, source coding techniques should account for both the mean and the variance of the distortion.

D. Channel Mismatch

In order to compare the robustness of the MED and VAPOR approaches we analyze their sensitivity to channel mismatch. Channel mismatch arises when the actual probability of loss in the channel is different from the loss probability estimated at the transmitter. Consider the "Foreman" sequence coded at 30 fps with $R = 300$ kbps. In this experiment, both the MED and VAPOR approaches are optimized assuming that the probability of loss in the channel is $\rho = 0.05$. We then transmit the same encoded sequences over a channel with different probability of packet loss in order to measure their robustness to channel mismatch.

As seen in Fig. 11, at $\rho = 0.05$, both approaches have roughly the same expected distortion. If the probability of loss in the channel is lower than expected, e.g., $\rho = 0.025$, then the MED approach has slightly lower expected distortion than VAPOR. This is because the MED approach has smaller source coding distortion. If the channel is worse than expected, the VAPOR approach shows significant resilience to channel mismatch as compared with MED. One explanation for this is that the VAPOR approach inherently uses more intracoding then the MED approach, as shown in Fig. 8. Similar results have been obtained for other sequences and parameter settings.

In Fig. 11, the line labeled "matched MED" is the MED approach optimized for each ρ. Similarly, the curve labeled "matched VAPOR" is the VAPOR approach optimized for each ρ. It is interesting to note that the mismatched VAPOR approach (i.e., optimized for $\rho = 0.05$) performs relatively close to the

matched VAPOR and matched MED approaches over a large range of loss probabilities. On the other hand, the difference between the mismatched MED and the matched MED approaches increases more rapidly as the channel mismatch increases. Thus, it may be more critical to know the actual probability of loss in the channel, when using a MED approach. By using a variance-aware resource allocation technique, one can be more confident that the distortion will not increase as dramatically if the probability of loss is higher than expected.

VII. Conclusion

This paper identifies the variance of the end-to-end distortion as an important quantity for characterizing video quality in packet lossy networks. A major contribution is the added flexibility and capability to control both the expected value and the variance of the distortion. As shown through experimental results, the proposed approach helps limit error propagation and is more robust to channel mismatch than approaches which only account for the expected distortion.

Although resilient video coding has been the primary focus, the concepts introduced in this paper are applicable to other coding and transmission systems. For example, in [12], variance-aware resource allocation has been considered in the context of joint source coding and transmission power adaptation. In addition, the ideas presented here can be utilized in other cost-distortion optimization frameworks [30], such as joint source-channel coding, and video over DiffServ networks [31].

The distortion estimation techniques as well as the resource allocation schemes developed here are optimal and, thus, computationally intensive. Developing low complexity algorithms based on the concepts introduced here is an important area of future work. Understanding human sensitivity to the different spatio-temporal artifacts caused by source and channel distortion is another area that requires significant research. This understanding will help determine the perceptual importance of the mean and the variance of the end-to-end distortion in video communication systems.

References

[1] R. O. Hinds, "Robust Mode Selection for Block-Motion-Compensated Video Encoding," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, MA, 1999.

[2] 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. Sel. Areas Commun.*, vol. 18, no. 6, pp. 977–995, Jun. 2000.

[3] 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.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.

[4] 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. 2003.

[5] H. Yang and K. Rose, "Recursive end-to-end distortion estimation with model-based cross-correlation approximation," presented at the Int. Conf. Image Processing, Barcelona, Spain, Sep. 2003.

[6] 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, Jun. 2000.

[7] 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, Jun. 2000.

[8] T. Stockhammer, T. Wiegand, and S. Wenger, "Optimized transmission of H.26L/JVT coded video over packet-lossy networks," presented at the Int. Conf. Image Processing, Rochester, NY, 2002.

[9] 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. 4, pp. 511–523, Jun. 2002.

[10] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Trans. Multimedia*, to be published.

[11] 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. 4, pp. 411–424, Jun. 2002.

[12] 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 Int. Conf. Image Processing, Barcelona, Spain, 2003.

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

[14] 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. 8, pp. 361–372, Dec. 1992.

[15] Y. Sermadevi, M. Masry, and S. S. Hemami, "MINMAX rate control with a perceived distortion metric," presented at the SPIE Visual Communication Image Processing, San Jose, CA, Jan. 2004.

[16] 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," presented at the Int. Conf. Image Processing, Rochester, NY, 2002.

[17] 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 Int. Conf. Image Processing, Barcelona, Spain, Sep. 2003.

[18] 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.

[19] "Video Coding for Low Bitrate Communications, ITU-T," ITU, ITU-T Recommendation H.263 Version 2, 1997.

[20] "Coding of Audio-Visual Objects—Part2: Visual," ISO/IEC 14 496-2 (MPEG-4 Visual Version 1), 1999.

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

[22] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 11, pp. 23–50, Nov. 1998.

[23] D. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.

[24] Y. Wang, G. Wen, S. Wenger, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Process. Mag.*, vol. 17, no. 7, pp. 61–82, Jul. 2000.

[25] S. Wenger, "H.264/AVC over IP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 645–656, Jul. 2003.

[26] L. Ozarow, S. Shamai, and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 359–378, May 1994.

[27] S. Nanda, K. Balachandran, and S. Kumar, "Adaption techniques in wireless packet data services," *IEEE Commun. Mag.*, vol. 38, no. 1, pp. 54–64, Jan. 2000.

[28] E. Soyak, Y. Eisenberg, F. Zhai, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Channel modeling and its effects on the end-to-end distortion in wireless video communications," presented at the Int. Conf. Image Processing, Singapore, 2004.

[29] R. Zhang, S. L. Regunathan, and K. Rose, "Prescient mode selection for robust video coding," in *Proc. Int. Conf. Image Processing*, Thessaloniki, Greece, 2001, pp. 974–977.

[30] A. K. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry, and T. N. Pappas, "Advances in efficient resource allocation for packet-based real-time video transmission," *Proc. IEEE*, vol. 93, no. 1, pp. 135–147, Jan. 2005.

[31] F. Zhai, C. E. Luna, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and packet classification for video streaming over differentiated services networks," *IEEE Trans. Multimedia*, vol. 7, no. 4, pp. 716–726, Aug. 2005.

Yiftach Eisenberg (S'02–M'04) received the B.S. degree in electrical engineering from the University of Illinois, Urbana-Champaign, in 1999, and the M.S. and Ph.D. degrees in electrical engineering from Northwestern University, Evanston, IL, in 2001 and 2004, respectively.

He is currently a Senior Engineer at BAE Systems, Advanced Systems and Technology, Merrimack, NH. In 2000, he was a Visiting Researcher with Motorola Labs, Schaumburg, IL, in the Multimedia Research Laboratory. His primary research interests include signal processing, wireless communications, networking, and video compression and analysis.

Fan Zhai (S'99–M'04) received the B.S. and M.S. degrees in electrical engineering from Nanjing University, Nanjing, China, in 1996 and 1998, respectively, and the Ph.D. degree in electrical and computer engineering from Northwestern University, Evanston, IL, in 2004.

He is currently a System Engineer in the Digital Video Department, Texas Instruments, Dallas, TX. His primary research interests include image and video signal processing and compression, multimedia communications and networking, and multimedia analysis.

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, 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 September 1999, he joined the Department of Electrical and Computer Engineering, 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 IPSN'04, and, since 1997, he has been Co-Chair of the SPIE/IS&T Conference on Human Vision and Electronic Imaging. He was also Co-Chair of the 2005 IS&T/SPIE Symposium on Electronic Imaging: Science and Technology.

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 with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL. In 1998, he was on the technical staff at the MIT 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 National Science Foundation CAREER award.

Aggelos K. Katsaggelos (S'80–M'85–SM'92–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 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 a Professor, holding the Ameritech Chair of Information Technology. He is also the Director of the Motorola Center for Communications and a member of the Academic Affiliate Staff, Department of Medicine, at Evanston Hospital. He is the editor of *Digital Image Restoration* (New York: Springer-Verlag, 1991), coauthor 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), and the co-inventor of eight international patents

Dr. Katsaggelos is a member of the Publication Board of the IEEE PROCEEDINGS, the IEEE Technical Committees on Visual Signal Processing and Communications, and Multimedia Signal Processing, the Editorial Board of Academic Press, Marcel Dekker: Signal Processing Series, *Applied Signal Processing*, and *Computer Journal*. He has served as Editor-in-Chief of the *IEEE Signal Processing Magazine* (1997–2002), member of the Publication Boards of the IEEE Signal Processing Society, the IEEE TAB Magazine Committee, Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1990–1992), Area Editor for the journal *Graphical Models and Image Processing* (1992–1995), member of the Steering Committees of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1992–1997) and the IEEE TRANSACTIONS ON MEDICAL IMAGING (1990–1999), member of the IEEE Technical Committee on Image and Multi-Dimensional Signal Processing (1992–1998), and a member of the Board of Governors of the IEEE Signal Processing Society (1999–2001). 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).