OPTIMAL ERASURE PROTECTION FOR SCALABLY COMPRESSED VIDEO STREAMS WITH LIMITED RETRANSMISSION ON CHANNELS WITH IID AND BURSTY LOSS CHARACTERISTICS

Johnson Thie and David Taubman

The University of New South Wales, Sydney, Australia

ABSTRACT

We show how the Priority Encoding Transmission (PET) framework may be leveraged to exploit both unequal error protection and limited retransmission, for RD optimized delivery of streaming media. Previous work on scalable media protection with PET has largely ignored the possibility of retransmission. We focus on sources which can be modeled as independently compressed frames, where each element of each frame can be transmitted in one or both of two transmission slots. An optimization algorithm determines the level of protection for each element in each slot, subject to transmission bandwidth constraints. To balance the protection assigned to elements being transmitted for the first time with those being retransmitted, the proposed algorithm formulates a collection of hypotheses concerning its own behaviour in future transmission slots. Experimental results are reported using both IID and GE channel models, with a Motion JPEG2000 video source, demonstrating substantial performance benefits from the proposed framework.

1. INTRODUCTION

This paper addresses the robust transmission of scalable data through lossy communication channels, for applications in which limited retransmission is possible. We restrict our attention here to video streams which can be modeled as a sequence of independently compressed "source frames," $\mathcal{F}[n]$, each with its own collection of embedded elements, $\mathcal{E}_q[n]$.

Forward Error Correction (FEC) codes are often adopted for the transmission of real-time compressed data, based on the assumption that retransmission might cause unacceptable delays or violate delivery time constraints. In many applications, however, delivery time constraints merely serve to limit the number of round trips, and hence the number of existing retransmission opportunities. This perspective is adopted by several works, including [1]. The combination of both limited retransmission and FEC has also been considered in a variety of settings, including [2, 3].

The lossy communication channel considered in this work is that of a packet-based "erasure channel," in which each packet either arrives intact or is entirely lost. A key property of erasure channels is that the receiver knows exactly which packets have been lost (the "erasures"). In the context of erasure channels, Albanese et al.[4] introduced PET. The PET scheme works with a family of channel codes, all of which have the same codeword length, N, but different source lengths, $1 \leq k \leq N$. We consider only "maximum distance separable" (MDS) codes, which have the key property that receipt of any k out of the N symbols in a (N, k) codeword is sufficient to recover the k source symbols. The amount of redundancy, $R_{N,k} = N/k$, determines the strength of the code. It is convenient to augment the set of channel codes with the special value $k = \infty$, for which $R_{N,\infty} = 0$, meaning that the element is not transmitted at all. Given a collection of source elements, $\mathcal{E}_1, \ldots, \mathcal{E}_Q$, having uncoded lengths, $L_1, \ldots, L_Q$, and channel code redundancies, $R_{N,k_1} \geq \cdots \geq R_{N,k_Q}$, the PET scheme packages the source elements into N network packets, which we call a "PET frame." This packaging has the key property that whenever sufficient packets are received to decode element $\mathcal{E}_q$, elements $\mathcal{E}_1$ through $\mathcal{E}_{q-1}$ can also be successfully decoded.

As mentioned, we are interested in applications where limited retransmission of lost data is possible. To simplify matters, each scalable source frame $\mathcal{F}[n]$ is assigned only two opportunities for transmission: a primary "transmission slot" $\mathcal{T}[n]$; and a secondary transmission slot $\mathcal{T}[n+\kappa]$, during which lost information may be retransmitted. The retransmission delay, κ, is chosen so that this separation is long enough for the transmitter to discover whether or not packets sent during the primary transmission slot arrived successfully. Figure 1 illustrates the relationship between source frames and transmission slots. We assume that the delivery time constraints is always sufficiently large to allow one retransmission[1].

During transmission slot $\mathcal{T}[n]$, the transmitter must determine how best to distribute the available bandwidth between the primary transmission of $\mathcal{F}[n]$ and the secondary transmission (retransmission) of $\mathcal{F}[n-\kappa]$, without knowing how many packets will be lost or what protection might be applied to the retransmission of any lost data from frame $\mathcal{F}[n]$ in its secondary transmission slot $\mathcal{T}[n+\kappa]$. To this end, we formulate a collection of hypotheses concerning the current frame's secondary transmission, which will take place in the future. Each hypothesis corresponds to a separate possible outcome regarding the number of transmitted packets which might be lost. Using these hypotheses and their likelihoods, decisions are made in such a way as to maximize a global Lagrangian cost function over the entire video sequence, based on the information at hand in slot $\mathcal{T}[n]$, while satisfying a total slot length constraint, $L_{\max}$. A closely related optimization objective is proposed in [1], although our approach benefits tremendously from the PET paradigm. We use the term LR-PET (Limited Retransmission PET) to describe this novel framework.

A full development of the LR-PET framework may be found in [5]. The present paper provides experimental results comparing LR-PET with other PET-based schemes. Most notably, we demonstrate the importance of including hypotheses regarding future re-

[1]We may assume that the receiver always have sufficiently large buffer to store the video frames prior to processing/display, so that them can be transmitted early in advance.

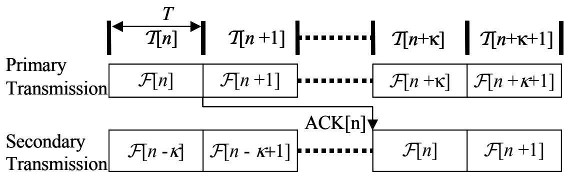

Fig. 1. *Relationship between source frames $\mathcal{F}[n]$ and transmission slots $\mathcal{T}[n]$.* ACK$[n]$ *signifies acknowledgment information, used to inform the transmitter of packet losses.*

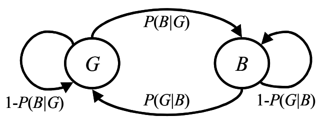

Fig. 2. *Simplifed Gilbert channel model.*

transmission within the optimization framework. This element is notably missing from the only other scheme we know of which considers both PET and retransmission [3]. Our experiments employ both IID and Gilbert-Elliott based packet loss models.

2. CHANNEL MODEL

The channel model we use is that of an erasure channel, where the receiver knows exactly which packets have been lost. We consider both IID and the "simplified Gilbert" (SG) [6] packet loss models[2]. With the IID model, the probability of receiving at least k out of N packets with no error is $P_{N,k} = \sum_{i=k}^{N} \binom{N}{i} (1 - P_E)^i P_E^{N-i}$ where P_E is the probability of packet erasure.

The SG model is a 2-state Markov model of the channel condition, as shown in Figure 2. The "good" state G and "bad" state B correspond to packets arriving with no error and packets which are entirely lost, respectively. The SG model is determined completely by P_E and the average erasure burst length L_E, since

$$P_E = \frac{P(B|G)}{P(B|G) + P(G|B)} \quad \text{and} \quad L_E = \frac{1}{P(G|B)}$$

The probability of receiving at least k out of N packets with no error, $P_{N,k}$, is equivalent to the probability of visiting state G at least k times in N steps. We may write this as $P_{N,k} = \sum_{i=k}^{N} \gamma_G(N,i) + \gamma_B(N,i)$, where $\gamma_G(N,k) = \sum_{m=\{G,B\}} \gamma_m(N-1,k-1) P(G|m)$ denotes the probability of visiting state G exactly k times in N steps, ending in G, while $\gamma_B(N,k) = \sum_{m=\{G,B\}} \gamma_m(N-1,k) P(B|m)$ denote the probability of visiting state G exactly k times in N steps, ending in state B. These expressions can be evaluated recursively with initial conditions, $\gamma_G(1,1) = 1 - P_E$, $\gamma_B(1,0) = P_E$ and $\gamma_B(1,1) = \gamma_G(1,0) = 0$.

It is convenient to parametrize $P_{N,k}$ and $R_{N,k}$ by a single parameter r (the "redundancy index"), writing

$$r = \begin{cases} N+1-k, & k \neq \infty \\ 0, & k = \infty \end{cases}, \quad P(r) = P_{N,k} \text{ and } R(r) = R_{N,k}$$

[2]The simplified Gilbert model matches packet-based erasure channels and can be readily presented in terms of the parameters of the more general Gilbert-Elliott channel which operates at the bit/symbol level [7].

We also write $k_{\min}(r) = N + 1 - r$ for the minimum number of packets which must be received if an element with redundancy index r is to be recovered.

3. PET WITHOUT RETRANSMISSION

We briefly review the problem of assigning optimal redundancy indices to elements $\mathcal{E}_1$ through $\mathcal{E}_Q$ of a single source frame, subject to the following constraint on the length of the PET frame.

$$L = \sum_{q=1,\ldots,Q} L_q R(r_q) \leq L_{\max}$$

When retransmission is not possible, we use this method to protect each frame $\mathcal{F}[n]$ within its own transmission slot $\mathcal{T}[n]$.

We assume that the source elements exhibit a simple dependency structure, $\mathcal{E}_1 \prec \cdots \prec \mathcal{E}_Q$, meaning that elements $\mathcal{E}_1$ through $\mathcal{E}_q$ must all be available before $\mathcal{E}_{q+1}$ can be correctly decoded. The PET optimization objective is an expected utility

$$U = U_0 + \sum_{q=1,\ldots,Q} U_q P(r_q) \qquad (1)$$

where U_0 is the amount of utility at the receiver when no source element is received and U_q is the utility associated with receiving $\mathcal{E}_q$. Equation (1) holds so long as the underlying utility measure is additive and $r_1 \geq r_2 \geq \cdots \geq r_Q$, by virtue of the PET packaging property that $\mathcal{E}_1$ through $\mathcal{E}_q$ are correctly decoded whenever $\mathcal{E}_{q+1}$ is. U_q can then only be used after the element $\mathcal{E}_q$ assigned with a code redundancy index r_q can be recovered at the receiver. Commonly, $-U$ represents the Mean Squared Error (MSE) of the reconstructed video frame and U_q corresponds to a reduction in MSE associated with recovery of element $\mathcal{E}_q$.

For the purposes of this paper, we assume that the source utility-length characteristic is convex, meaning that $U_1/L_1 \geq \cdots \geq U_Q/L_Q$[3]. Several schemes (e.g. [8, 9]) have been reported for optimizing equation (1) under these conditions. Briefly, the constrained optimization problem may be converted to a family of unconstrained optimization problems, parametrized by a quantity $\lambda > 0$. Let U_λ and L_λ denote the expected utility and transmission length associated with the set of redundancy indices $\{r_{\lambda,q}\}$, which maximize the functional

$$J_\lambda = U_\lambda - \lambda L_\lambda = \sum_{q=1,\ldots,Q} \underbrace{U_q P(r_{\lambda,q}) - \lambda L_q R(r_{\lambda,q})}_{J_{\lambda,q}}$$

subject to $r_{\lambda,1} \geq r_{\lambda,2} \geq \cdots \geq r_{\lambda,Q}$. Evidently, it is impossible to increase U beyond U_λ, without also increasing L beyond L_λ. Thus, if we can find λ such that $L_\lambda = L_{\max}$, the set $\{r_{\lambda,q}\}$ is an optimal solution to our constrained problem. In practice, the discrete nature of the problem may prevent us from finding a value λ such that L_λ exactly equals $L_{\max}$, but if the elements are small enough, we are justified in ignoring this small source of suboptimality, selecting the smallest value of λ such that $L_\lambda \leq L_{\max}$.

If we temporarily ignore the constraint, $r_{\lambda,1} \geq \cdots \geq r_{\lambda,Q}$, the unconstrained optimization problem can be decomposed into a set of Q separate maximization problems, $J_{\lambda,q}$. This optimization problem arises in other contexts, such as the optimal truncation of embedded compressed bit-streams. The solutions $r_{\lambda,q}$ must belong to the set $\mathcal{H}_C$ which describes the convex hull of the $P(r)$ vs.

[3]This assumption arises from the Rate-Distortion characteristics of the video source coding, which is commonly modeled as a convex curve.

$R(r)$ characteristic. Let $0 = j_0 < j_1 < \cdots$ be an enumeration of the elements in $\mathcal{H}_C$, and $S_C(i) = \frac{P(j_i)-P(j_{i-1})}{R(j_i)-R(j_{i-1})}$, $i > 0$, be the "slopes" on the convex hull, where $\infty = S_C(0) \geq S_C(1) \geq \cdots$. The solution to our optimization problem can then be found from

$$r_{\lambda,q} = \max\{j_i \in \mathcal{H}_C \mid S_C(i) \geq \lambda L_q/U_q\}$$

Moreover, since the source is convex, we must have $\lambda L_q/U_q \leq \lambda L_{q+1}/U_{q+1}$, from which we deduce that $r_{\lambda,q} \geq r_{\lambda,q+1}$. Thus, the constraint $r_{\lambda,1} \geq r_{\lambda,2} \geq \cdots \geq r_{\lambda,Q}$ is satisfied after all.

4. THE LR-PET FRAMEWORK

This section provides a brief description of the LR-PET framework. A more rigorous development may be found in [5]. During each transmission slot, $\mathcal{T}[n]$, the transmitter selects channel code redundancy indices $r_q[n]$ for each element of the primary frame $\mathcal{F}[n]$, as well as retransmission redundancy indices $s_q[n-\kappa]$ for each element of frame $\mathcal{F}[n-\kappa]$, subject to a limit $L_{\max}$ on the total number of symbols that can be transmitted within any slot. Our objective is to jointly optimize the parameters $r_q[n]$ and $s_q[n-\kappa]$, subject to the length constraint $L_{\max}$ on slot $\mathcal{T}[n]$, bearing in mind that elements of $\mathcal{F}[n]$ will have the opportunity to be retransmitted in slot $\mathcal{T}[n+\kappa]$. We formulate hypotheses concerning such retransmission, writing $s_q^k[n]$ for the hypothetical redundancy index to be assigned to $\mathcal{E}_q[n]$ in transmission slot $\mathcal{T}[n+\kappa]$, in the event that only k of the N packets from slot $\mathcal{T}[n]$ are received.

The expected utility of frame $\mathcal{F}[n]$ may then be expressed as

$$E\left[U[n]\right] = U_0[n] + \sum_{k=0,\ldots,N} \rho_k \sum_q U_q[n]\, P\left(k, r_q[n], s_q^k[n]\right) \tag{2}$$

where $\rho_k = P_{N,k} - P_{N,k-1}$ is the probability that exactly k packets are received and

$$P\left(k, r_q, s_q^k\right) = \begin{cases} P\left(s_q^k\right) & , k < k_{\min}(r_q) \\ 1 & , k \geq k_{\min}(r_q) \end{cases}$$

is the probability that element $\mathcal{E}_q[n]$ will be recovered if it is assigned a redundancy index of $r_q[n]$ in $\mathcal{T}[n]$, if only k of the N packets transmitted in $\mathcal{T}[n]$ are received, and if the remaining data is retransmitted with redundancy index $s_q^k[n]$ in $\mathcal{T}[n+\kappa]$.

The expected total transmission length associated with frame $\mathcal{F}[n]$ may be expressed as

$$E\left[L[n]\right] = \sum_{k=0,\ldots,N} \rho_k \sum_q L_q[n]\, R\left(k, r_q[n], s_q^k[n]\right) \tag{3}$$

where $R\left(k, r_q, s_q^k\right)$ is the ratio between the total coded length and the uncoded length of $\mathcal{E}_q[n]$, under the same conditions stated above for $P\left(k, r_q, s_q^k\right)$. It can be shown that

$$R\left(k, r_q, s_q^k\right) = R(r_q) + \begin{cases} \theta_{k,r_q} R\left(s_q^k\right) & , k < k_{\min}(r_q) \\ 0 & , k \geq k_{\min}(r_q) \end{cases}$$

where $\theta_{k,r_q} = 1 - k/k_{\min}(r_q)$ is the fraction of $\mathcal{E}_q[n]$'s source symbols which must be successfully retransmitted to allow complete recovery from the information in its primary and secondary transmission slots.

The idea behind LR-PET is to choose the $r_q[n]$, $s_q[n-\kappa]$ and $s_q^k[n]$ parameters in $\mathcal{T}[n]$ so as to maximize the global utility-length functional

$$E\left[\sum_n U[n]\right] - \lambda E\left[\sum_n L[n]\right] \tag{4}$$

adjusting λ to ensure that total encoded length in $\mathcal{T}[n]$ is no more than $L_{\max}$; note that this length depends on $r_q[n]$ and $s_q[n]$, while the global utility-length objective depends also on the hypothetical indices $s_q^k[n]$. Eliminating all terms from equation (4) which do not depend on $r_q[n]$, $s_q[n-\kappa]$ and $s_q^k[n]$, we find that our goal in $\mathcal{T}[n]$ is to maximize

$$J_\lambda\left(\left\{r_q, \mathbf{s}_q, s_q'\right\}_q\right) = \Psi_\lambda\left(\left\{r_q, \mathbf{s}_q\right\}_q\right) + J_\lambda'\left(\left\{s_q'\right\}_q\right) \tag{5}$$

subject to the constraints

$$r_q \geq r_{q+1},\ s_q^k \geq s_{q+1}^k,\ \text{and } s_q' \geq s_{q+1}',\ \forall k, q \tag{6}$$

where $\mathbf{s}_q \equiv \left(s_q^0[n], \ldots, s_q^N[n]\right)$ is the vector of hypothetical redundancy indices, $r_q \equiv r_q[n]$, $s_q' \equiv s_q[n-\kappa]$, and

$$\Psi_\lambda\left(\left\{r_q, \mathbf{s}_q\right\}_q\right) = \sum_{k=0,\ldots,N} \rho_k \sum_q U_q P\left(k, r_q, s_q^k\right) - \lambda \sum_{k=0,\ldots,N} \rho_k \sum L_q R\left(k, r_q, s_q^k\right)$$

$$J_\lambda'\left(\left\{s_q'\right\}_q\right) = \sum_q \left[U_q' P\left(k', r_q', s_q'\right) - \lambda L_q' R\left(k', r_q', s_q'\right)\right]$$

We are using primes here to denote quantities associated with the earlier frame $\mathcal{F}[n-\kappa]$ and note that we already know the values of $r_q' = r_q[n-\kappa]$ and k', the number of packets received from $\mathcal{T}[n-\kappa]$. The constraints in equation (6) ensure the validity of the expected utility expression in equation (2). They may be imposed without loss of generality, since there is always at least one solution which maximizes equation (4) and satisfies equations (6).

It is not hard to see that $J_\lambda'\left(\left\{s_q'\right\}_q\right)$ can be maximized separately, following exactly the same procedure outlined for regular PET in Section 3. The optimization of $\Psi_\lambda\left(\left\{r_q, \mathbf{s}_q\right\}_q\right)$ is more complex, but not so complex as one might expect. To shed light on this, we summarize here the main results of [5]:

1. Writing $\Psi_\lambda\left(\left\{r_q, \mathbf{s}_q\right\}_q\right) = \sum_q \Psi_{\lambda,q}(r, \mathbf{s}_q)$, it turns out that the $\Psi_{\lambda,q}(r, \mathbf{s}_q)$ can be maximized independently for each q; the resulting solutions will then automatically satisfy the constraints of equation (6), subject to some reasonable assumptions. This is the same property which makes the PET optimization problem of Section 3 so attractive, although the result is not nearly so obvious in the case of LR-PET.
2. To maximize a single term $\Psi_{\lambda,q}(r, \mathbf{s}_q)$ requires at most $2N^2$ relatively simple comparisons between scaled utility-length slopes.
3. For larger Q, a divide and conquer approach allows all Q terms $\Psi_{\lambda,q}(r, \mathbf{s}_q)$ to be maximized with at most $NQ + 2N^2 \log_2(Q+1)$ such comparisons.

5. EXPERIMENTAL RESULTS

We present experimental results to compare the performance between LR-PET with transmission delay κ, and two alternative protection schemes based on PET, using both the IID and the GE

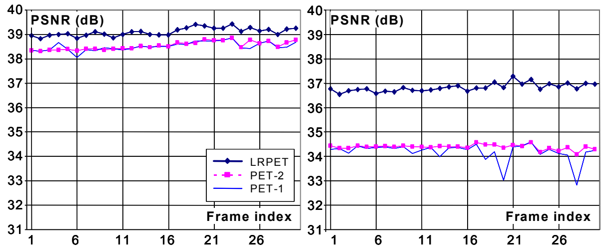

Fig. 3. *Comparative results with IID packet loss model. Left:* $P_E = 0.01$. *Right:* $P_E = 0.2$.

packet loss models. The first case, which we call PET-1, protects the scalable source frames using the technique described in Section 3 without any retransmission. The second case, PET-2, is similar to LR-PET, except that we deliberately omit the retransmission hypotheses from equations (2) and (3); equivalently, we fix $s_q^k[n] = 0, \forall\, k, q, n$. In this case, the optimization problem has essentially the same complexity as PET, except that there are more elements to be considered in the optimization, since retransmission is allowed.

We use a Motion JPEG2000 compressed video sequence, consisting of 30 monochrome progressive-scan frames, each measuring 720×1280 and compressed with 15 precincts, 12 quality layers, for a total of $Q = 180$ elements, $\mathcal{E}_q[n]$. The video sequence name is "bigships." The packet erasure probabilities for both IID and SG models are 0.01 and 0.2. The erasure burst lengths, L_E, for the SG model are 2 and 20 packets. We use negative MSE as our measure of utility, taking averages over a large number of experiments and reporting results in terms of PSNR. Although our source consists of only 30 frames, we cycle through these frames many times, effectively creating a much larger sequence.

Figure 3 compares LR-PET, PET-1 and PET-2 with the IID channel model. These results clearly show the benefits of including retransmission hypotheses in the optimization objective (LR-PET vs. PET-2), with improvements of more than 2 dB at higher loss rates. This is important, since the only other work we know of which combines PET with limited retransmission [3] does not include such hypotheses. Our observation that PET-2 provides little benefit over PET-1 is consistent with theirs, although there are some differences in the problem formulation.

Figure 4 compares LR-PET, PET-1 and PET-2 using the SG channel model. The results show that LR-PET also generally performs best. In particular, the PSNR is improved by up to 4 dB relative to PET-2 and up to 6 dB relative to PET-1. The results also show that large erasure burst lengths not only decreases the overall PSNR across all frames but also increases the PSNR variation from frame-to-frame.

6. CONCLUSIONS

We present the LR-PET framework which assigns protection to source elements using hypotheses regarding future retransmission. It is important to employ the PET framework so that the optimization problem can be solved independently, thereby reducing the algorithm complexity. The experimental results reveal performance advantages from the inclusion of limited retransmission and its hypotheses.

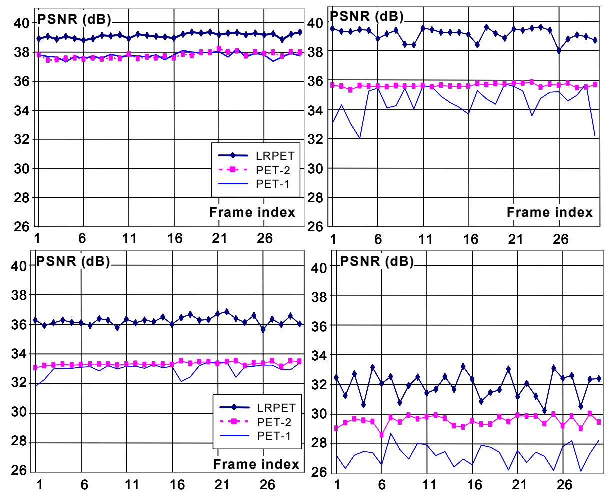

Fig. 4. *Comparative results with SG packet loss model. Top-left:* $P_E = 0.01$ *and* $L_E = 2$. *Top-right:* $P_E = 0.01$ *and* $L_E = 20$. *Bottom-left:* $P_E = 0.2$ *and* $L_E = 2$. *Bottom-right:* $P_E = 0.2$ *and* $L_E = 20$.

7. REFERENCES

[1] P. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *Microsoft Tech. Report MSR-TR-2001-35 (submitted to IEEE Trans. Multimedia)*, February 2001.

[2] P. Chou, A. Mohr, A. Wang, and S. Mehrotra, "FEC and pseudo-ARQ for receiver-driven layered multicase of audio and video," *Proc. IEEE Data Compression Conf. (Snowbird)*, pp. 440–449.

[3] T. Gan and K.-K. Ma, "Sliding window packetization for forward error correction based multiple description transcoding," *Proc. Int. Conf. Acoust. Speech and Sig. Proc.*, vol. 5, pp. 756–759.

[4] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1737–1744, November 1996.

[5] D. Taubman and J. Thie, "Optimal erasure protection for scalably compressed video streams with limited retransmission," *To appear in the IEEE Trans. on Image Processing*, 2004.

[6] J. Yee and E. W. Jr., "Evaluation of the performance of error-correcting codes on a gilbert channel," *IEEE Trans. Commun.*, vol. 43, pp. 2316–2323, August 1995.

[7] M. Zorzi and R. Rao, "On the statistics of block errors in bursty channels," *IEEE Trans. Commun.*, vol. 45, pp. 660–667, June 1997.

[8] R. Puri and K. Ramchandran, "Multiple description source coding using forward error correction codes," vol. 1, pp. 342–346, 1999.

[9] A. Mohr, E. Riskin, and R. Ladner, "Unequal loss protection: Graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 819–828, June 2000.