A SPECTRAL APPROACH TO RECURSIVE END-TO-END DISTORTION ESTIMATION FOR SUB-PIXEL MOTION-COMPENSATED VIDEO CODING

Jingning Han, Vinay Melkote, and Kenneth Rose*

Department of Electrical and Computer Engineering, University of California Santa Barbara
{jingning,melkote,rose}@ece.ucsb.edu

ABSTRACT

Error resilient video coding critically relies on the accuracy of end-to-end distortion estimation. An established solution, the recursive optimal per-pixel estimate (ROPE), is based on tracking the first and second moments of the decoder reconstructed pixels. This paper is focused on an alternative estimation approach, the *spectral coefficient-wise optimal recursive estimate* (SCORE), whose recursion is performed in the transform domain. The SCORE formulation is extended to derive a new technique for effective end-to-end distortion estimation, which accounts for sub-pixel motion compensation. Specifically, this technique exploits properties of the transform, such as coefficient de-correlation and energy compaction, to overcome ROPE's remaining shortcoming due to the proliferation of cross-correlation terms requiring excessive complexity or relatively crude approximations. Experiments show that the accuracy of SCORE matches ROPE in the full-pixel motion compensation setting, where ROPE is known to be optimal. More importantly, in the problematic setting of sub-pixel motion compensation, SCORE substantially outperforms ROPE and yields highly accurate distortion estimation.

Index Terms— end-to-end distortion, joint source channel coding, sub-pixel motion compensation

1. INTRODUCTION

Motion compensated prediction is employed in most video coders to remove temporal redundancies, at the expense of increased vulnerability to packet loss, due to temporal and spatial error propagation via the prediction loop. Many error resilience tools and paradigms have been employed to mitigate this problem, including forward error correction, intra refresh, multiple description coding, macroblock retransmission, etc., (see e.g. [1] for an overview of relevant techniques). Since error resilience typically introduces redundancies in the compressed signal, and hence incurs additional bit-rate costs, the fundamental optimization problem that underlies the coder is formulated in terms of the trade-off between bit-rate and the distortion perceived at the decoder, also referred to as end-to-end distortion (EED). Clearly, optimization of encoding decisions depends directly on the encoder's ability to accurately estimate the EED, while accounting for all factors, including compression, packet loss and error propagation due to the prediction loop, and concealment at the decoder. The recursive optimal per-pixel estimate (ROPE) [2], which originated in our lab, is an efficient and effective approach to optimally estimate the EED. Since the packet losses are random, the encoder must treat the decoder reconstruction of a pixel as a random variable. The main idea of ROPE is to recursively calculate the first and second moments of reconstructed pixels, via update equations that explicitly account for motion compensated prediction, packet loss rate, and concealment at the decoder. The optimal EED estimate is then directly obtained from the first and second moments of the reconstructed pixels. ROPE has been successfully incorporated into various methods for error-resilient video coding, including for example [3, 4].

The basic version of ROPE [2] was extended in [3, 5] to better comply with current standards by accounting for sub-pixel motion compensation, which involves inter-pixel correlation terms. Most sequences show performance advantage when using sub-pixel motion compensated coding, which generates pixels seated on the sub-pixel grid as prediction reference, through interpolation filtering of the reconstructed frame [6]. Ideally ROPE can estimate the EED accurately by tracking the cross correlation between every pair of pixels inside the frame, in addition to the first and second moments of each individual pixel, but this incurs impractically large complexity/memory requirements. Several pixel domain model-based methods were proposed in [3, 5] to approximate or model the cross correlation term given marginal first and second moments.

An alternative perspective is provided by estimating the EED in transform domain. There are source coding approaches of significant interest that involve operations that are recursive in the transform domain. In particular, [7, 8, 9] propose estimation-theoretic approaches for video source encoding/decoding that offer substantial compression gains, by recursively operating in the transform domain, typically the discrete cosine transform (DCT). Specifically, these approaches view the sequence of DCT coefficients at a given spatial frequency, from blocks along a motion trajectory across consecutive frames, as an autoregressive (AR) process, and exploit this per-coefficient AR model to estimate the coefficients of a given block. The need to provide such applications with a ROPE-like EED estimate capable of accounting for error propagation due to recursive operations in the transform domain for effective error resilience motivates our recently proposed *spectral coefficient-wise optimal recursive estimate* (SCORE) [10]. It was derived in a general setting to recursively calculate the moments of each transform (in practice DCT) coefficient of blocks in a frame, and account for general transform domain operations. The efficacy of SCORE was shown in the context of the transform domain motion-compensated prediction scheme, when deployed over a lossy network [10]. Other notable DCT domain approaches include [11] where the motion compensation is restricted only on-grid blocks in previous frame and [12] where the recursion is performed in pixel domain as regular ROPE before converting to transform domain (for detailed analysis, see [10]). This paper substantially extends the SCORE framework in order to resolve the longstanding difficulty of ROPE with sub-pixel prediction. In this work, we further modify SCORE to better capture

*Vinay Melkote is now with Dolby Laboratories, Inc., 100 Potrero Avenue, San Francisco, CA 94103

inter-pixel correlation due to sub-pixel filtering. It is experimentally demonstrated that the proposed transform domain approach achieves higher estimation accuracy compared to the pixel domain model-based counterparts.

2. THE RECURSIVE OPTIMAL PER-PIXEL ESTIMATE

This section provides a brief review of ROPE and its various extensions to account for sub-pixel motion compensated coding. Consider point-to-point video communication, with encoder access to some statistical information about the network condition. For simplicity (but without implied loss of generality) assume that packet loss is statistically uniformly distributed, and let the packet loss rate (PLR), denoted p, be available to the encoder. Clearly, for optimal performance, the encoder must optimize its decisions with respect to the reconstructed video quality *at the decoder*. However, the decoder reconstruction is a random process as far as the encoder is concerned, with the ultimate effect of channel loss greatly complicated by error propagation through the prediction loop, error concealment efforts at the decoder, etc.

Let f_n^i denote the original value of pixel i in frame n, and let $\hat{f}_n^i$ denote its *encoder* reconstruction. The reconstructed value at the *decoder*, possibly after error concealment, is denoted by $\tilde{f}_n^i$, which is a random variable for the encoder. The overall expected distortion (in the mean squared sense) for this pixel is $E\{(f_n^i - \tilde{f}_n^i)^2\} = (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i\} + E\{(\tilde{f}_n^i)^2\}$. To evaluate this distortion only requires the first and second moments of the decoder reconstructed pixel $\tilde{f}_n^i$. ROPE employs the following recursion formulas, developed separately for the two cases of intra- and inter- coding, sequentially to compute these two moments for each pixel.

Intra-coding: The packet containing pixel i is received correctly with probability $1-p$, producing $\tilde{f}_n^i = \hat{f}_n^i$. If the packet is lost (with probability p), we conceal as $\tilde{f}_n^i = \tilde{f}_{n-1}^i$. The first and second moments of $\tilde{f}_n^i$ for an intra-coded pixel are computed as:

$$\begin{aligned} E\{\tilde{f}_n^i\}(I) &= (1-p)(\hat{f}_n^i) + pE\{\tilde{f}_{n-1}^i\}\,, \\ E\{(\tilde{f}_n^i)^2\}(I) &= (1-p)(\hat{f}_n^i)^2 + pE\{(\tilde{f}_{n-1}^i)^2\}\,. \end{aligned} \quad (1)$$

For simplicity we assume that all the intra-coded macroblocks are self-contained and serve as instantaneous refresh points if received.

Inter-coding: Let pixel i be predicted from pixel j in the previous frame, i.e., the encoder generates the prediction error $e_n^i = f_n^i - \hat{f}_{n-1}^j$, whose reconstruction (at encoder) is denoted by $\hat{e}_n^i$. Even if the current packet is correctly received, the decoder must use for prediction the *decoder's* reconstruction of pixel j in the previous frame, $\tilde{f}_{n-1}^j$, potentially different from $\hat{f}_n^j$. Thus the first and second moments of $\tilde{f}_n^i$ for an inter-coded pixel are:

$$\begin{aligned} E\{\tilde{f}_n^i\}(P) &= (1-p)(\hat{e}_n^i + E\{\tilde{f}_{n-1}^j\}) + pE\{\tilde{f}_{n-1}^i\}\,, \quad (2) \\ E\{(\tilde{f}_n^i)^2\}(P) &= (1-p)E\{(\hat{e}_n^i + \tilde{f}_{n-1}^j)^2\} + pE\{(\tilde{f}_{n-1}^i)^2\}. \end{aligned}$$

Once the first and second moments are calculated, the EED of the pixel is readily available. Employing ROPE to optimize inter/intra mode and quantization step selection within a rate-EED framework [2] has been demonstrated to provide substantial gains over heuristic methods for EED calculation.

The use of sub-pixel motion compensated coding requires linear interpolation of pixel values. For illustration, consider a simple example: $Z = (X+Y)/2$, where X and Y denote the reconstructed pixels, and Z the interpolated pixel, all at the decoder. The expectation of Z is computed directly from the first moments of X and Y: $E\{Z\} = \frac{1}{2}(E\{X\} + E\{Y\})$. However, the expression for the second moment $E\{Z^2\} = \frac{1}{4}(E\{X^2\} + E\{Y^2\} + 2E\{XY\})$ introduces a cross-correlation term $E\{XY\}$. Although these cross-correlation terms can be computed via additional recursions, an accurate estimation of all the required cross-correlation terms requires, in general, an order of N^2 additional computations (and memory units), where N is the number of pixels in a frame (see [5] for more details). Such complexity has been considered a practical limitation on the applicability of ROPE. Prior approaches address this issue by approximating the cross-correlation term in terms of the already computed marginal first and second moments of individual pixels.

Cauchy-Schwarz approximation Cauchy-Schwarz inequality suggests $E\{XY\} \leq \sqrt{E\{X^2\}E\{Y^2\}}$. It is argued in [3] that since the pixel values are always positive, the cross-correlation $E\{XY\}$ trends toward its upper bound, and can be approximated via equality.

Pixel distance model In [5] a pixel distance dependent correlation coefficient model is proposed. Note that the correlation coefficient between X and Y is

$$\rho_{XY} = (E\{XY\} - E\{X\}E\{Y\})/(\sigma_X \sigma_Y), \quad (3)$$

where σ_X and σ_Y denote the standard deviation of X and Y respectively. This is modeled as: $\rho_{XY} = e^{-\alpha d_{XY}}$, where d_{XY} is Euclidian distance between pixels X and Y, and α is a constant.

3. SPECTRAL COEFFICIENT-WISE OPTIMAL RECURSIVE ESTIMATE

Recently, in [10], we introduced SCORE for the purpose of providing a ROPE-like technique that works directly in transform domain, and is capable of capturing operations performed therein. Instead of calculating moments and distortion of individual pixels as ROPE does, SCORE tracks the moments and distortion of individual transform coefficients (we constrain the transform to be DCT with block size restricted to 4x4).

Let $x_n^{k,m}$ denote the uncoded value of transform coefficient m in block k of frame n, and $\hat{x}_n^{k,m}$ and $\tilde{x}_n^{k,m}$ the encoder and decoder reconstructions of this coefficient respectively. Note that this block may not be predicted from an on-grid reference block in the previous frame. Let $u_n^{k,m}$ denote the uncoded value of coefficient m in this (possibly off-grid) reference block.[1] Again the encoder and decoder reconstructions of this coefficient are denoted as $\hat{u}_n^{k,m}$ and $\tilde{u}_n^{k,m}$. The encoder considers $\tilde{x}_n^{k,m}$ and $\tilde{u}_n^{k,m}$ as random variables due to the stochastic nature of packet loss. The expected distortion at coefficient $x_n^{k,m}$, called $\delta_n^{k,m}$, is

$$E\{(x_n^{k,m} - \tilde{x}_n^{k,m})^2\} = (x_n^{k,m})^2 - 2x_n^{k,m}E\{\tilde{x}_n^{k,m}\} + E\{(\tilde{x}_n^{k,m})^2\}.$$

The computation of $\delta_n^{k,m}$ only requires the first and second moments of the decoder reconstruction $\tilde{x}_n^{k,m}$. SCORE employs the following recursion functions, developed separately for the two cases of intra- and inter-coding, to sequentially compute these two moments for each transform coefficient in a frame.

Intra-coding: The recursions are the same as in ROPE, albeit with transform coefficients replacing pixels. Since the assumed concealment is "slice copy", if $\hat{x}_n^{k,m}$ is unavailable due to packet loss, it is concealed as $\tilde{x}_{n-1}^{k,m}$, i.e., it is equivalent to copying in pixel domain.

$$\begin{aligned} E\{\tilde{x}_n^{k,m}\}(I) &= (1-p)(\hat{x}_n^{k,m}) + pE\{\tilde{x}_{n-1}^{k,m}\}\,, \\ E\{(\tilde{x}_n^{k,m})^2\}(I) &= (1-p)(\hat{x}_n^{k,m})^2 + pE\{(\tilde{x}_{n-1}^{k,m})^2\}. \end{aligned} \quad (4)$$

[1]Note that while $u_n^{k,m}$ is indexed by n and k to indicate the location on the current frame it provides a reference for, it is in fact a function of pixels in frame $n-1$.

Inter-coding: We define $\hat{y}_n^{k,m}$ to be the quantized transform coefficient residual. Following arguments similar to ROPE it can be shown that,

$$E\{\tilde{x}_n^{k,m}\}(P) = (1-p)(\hat{y}_n^{k,m} + E\{\tilde{u}_n^{k,m}\}) + pE\{\tilde{x}_{n-1}^{k,m}\},$$
$$E\{(\tilde{x}_n^{k,m})^2\}(P) = (1-p)((\hat{y}_n^{k,m})^2 + 2\hat{y}_n^{k,m}E\{\tilde{u}_n^{k,m}\} + E\{(\tilde{u}_n^{k,m})^2\}) + pE\{(\tilde{x}_{n-1}^{k,m})^2\}\,. \quad (5)$$

Note that these equations involve the first and second moments of transform coefficients of the motion compensated block, which is potentially off-grid. We thus propose a complementary method to extract the required moments of such blocks from the available moments of on-grid blocks in frame $n-1$.

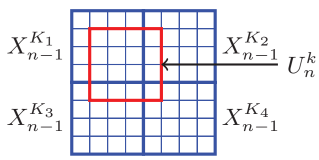

Fig. 1. Each off-grid block in a frame overlaps with 4 on-grid blocks. Here the blue blocks are on-grid, and the black off-grid block is employed for motion compensated prediction in the subsequent frame.

Any off-grid block in a frame overlaps with at most four on-grid blocks (Fig. 1). Let block U_n^k shown in the figure be the reference block for the current block k in frame n. This block, located in frame $n-1$, overlaps with on-grid blocks $X_{n-1}^{k_i}$ in the frame. The decoder reconstruction of block U_n^k is associated with coefficients $\tilde{u}_n^{k,m}$. Since DCT is a linear transformation, there exist constants $a_{i,m}$, named *construction constants*, such that,

$$\tilde{u}_n^{k,m} = \sum_{i=1}^{4}\sum_{m=0}^{15} a_{i,m}\tilde{x}_{n-1}^{k_i,m}\,. \quad (6)$$

These constants purely depend on the position of U_n^k relative to the on-grid blocks. Thus, the first moment of $u_n^{k,m}$ is simply

$$E\{\tilde{u}_n^{k,m}\} = \sum_{i=1}^{4}\sum_{m=0}^{15} a_{i,m}E\{\tilde{x}_{n-1}^{k_i,m}\}\,. \quad (7)$$

The second moment of $u_n^{k,m}$ is more complicated, and involves cross-correlations of DCT coefficient pairs of the on-grid blocks:

$$E\{(\tilde{u}_n^{k,m})^2\} = \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{m=0}^{15}\sum_{l=0}^{15} a_{i,m}a_{j,l}E\{\tilde{x}_{n-1}^{k_i,m}\tilde{x}_{n-1}^{k_j,l}\}\,. \quad (8)$$

The advantage of transform domain is that it largely decorrelates the block. Specifically, as we shall see later in the results, for the case of full-pel motion compensated coding the following assumption of 'uncorrelatedness' holds well in the DCT domain:

$$E\{\tilde{x}_n^{k_i,m}\tilde{x}_n^{k_j,l}\} \approx E\{\tilde{x}_n^{k_i,m}\}E\{\tilde{x}_n^{k_j,l}\} \text{ when } j \neq i \text{ or } l \neq m. \quad (9)$$

4. EXTENSION OF SCORE FOR SUB-PIXEL MOTION COMPENSATED CODING

The main contribution of this paper is a significant extension of SCORE to effectively provide a ROPE-like technique for sub-pixel motion compensated coding. Unlike in the full-pixel case where each pixel is predicted from a reconstructed pixel in the reference frame, in the sub-pixel case in effect the reference for any pixel in the current frame is an area of pixels (typically of dimension 6x6) in the previous frame (on account of the interpolation involved), and thereby its moments depend on the joint statistics of all the pixels in this area. This pixel domain operation consequently necessitates two modifications to the SCORE approach.

Construction constants

Let us say that the reference block U_n^k is located on the sub-pixel grid. Generating the 4x4 block U_n^k via a 6-tap filter could potentially involve as many as 9x9 reconstructed pixels in frame $n-1$, and hence involve as many as 9 on-grid blocks (12x12 pixels) of that frame. Since interpolation and DCT are both linear operations, there exist a new set of construction constants $b_{i,m}$, such that,

$$\tilde{u}_n^{k,m} = \sum_{i=1}^{9}\sum_{m=0}^{15} b_{i,m}\tilde{x}_{n-1}^{k_i,m}\,. \quad (10)$$

Consequently $b_{k,m}$ will be used to compute the moments of $\tilde{u}_n^{k,m}$ akin to (7) and (8).

Recursion approximations

Note that (9) is only an approximation. Although it holds well in the full-pixel case (as will be evident from Sec. 5), experiments revealed that the resulting approximation error propagates more aggressively in the case of sub-pixel motion compensated coding (due to the mixing of these errors in the pixel domain via sub-pixel interpolation filtering). This necessitates the following modifications to the approximation in (9).

Cross-correlation within a block: Consider two transform coefficients $\tilde{x}_{n-1}^{k,i}$ and $\tilde{x}_{n-1}^{k,j}$ that are inside the *same* reference block k but at different frequencies. Let $\tilde{x}_r^i$ denote the decoder reconstruction of $x_{n-1}^{k,i}$ when the packet containing the block is received, and $\tilde{x}_e^i$ denote the reconstruction when it is lost (i.e., after error concealment). The notation $\tilde{x}_r^j$ and $\tilde{x}_e^j$ is to be interpreted similarly with respect to $\tilde{x}_{n-1}^{k,j}$. Note that $\tilde{x}_r^i$ and $\tilde{x}_e^i$, are both random variables with regards to the encoder, and their first moments can be accurately tracked as

$$E\{\tilde{x}_r^i\} = E\{\tilde{u}_{n-1}^{k,i}\} + \hat{y}_{n-1}^{k,i}$$
$$E\{\tilde{x}_e^i\} = E\{\tilde{x}_{n-2}^{k,i}\}\,. \quad (11)$$

Since all the transform coefficients of a block are contained in a single packet, they are received or lost simultaneously. Thus, the cross-correlation of $\tilde{x}_{n-1}^{k,i}$ and $\tilde{x}_{n-1}^{k,j}$ is *exactly*:

$$E\{\tilde{x}_{n-1}^{k,i}\tilde{x}_{n-1}^{k,j}\} = (1-p)E\{\tilde{x}_r^i\tilde{x}_r^j\} + pE\{\tilde{x}_e^i\tilde{x}_e^j\}. \quad (12)$$

This involves the knowledge of the cross correlations $E\{\tilde{x}_r^i\tilde{x}_r^j\}$ and $E\{\tilde{x}_e^i\tilde{x}_e^j\}$. We now appeal to the 'uncorrelatedness' assumption in DCT domain. Specifically:

$$E\{\tilde{x}_{n-1}^{k,i}\tilde{x}_{n-1}^{k,j}\} \approx (1-p)E\{\tilde{x}_r^i\}E\{\tilde{x}_r^j\} + pE\{\tilde{x}_e^i\}E\{\tilde{x}_e^j\}. \quad (13)$$

In other words, uncorrelatedness is treated separately for the concealment case, and the case when the packet is received.

Inter-block correlation: In the case of cross-correlation terms that involve coefficients from two different blocks there is no guarantee that both coefficients will be lost or received simultaneously. Although an extension of (13) for this scenario might still be feasible, in the current paper we follow a simple alternative: due to the energy compaction property of DCT, the dominant inter-block cross correlation term would likely be that between DC components, and the corresponding correlation is assumed to be unity. While more careful modeling will further improve the estimate accuracy, it is experimentally demonstrated (see Sec. 5) that this rough approximation already outperforms pixel domain approaches.

5. RESULTS

We first compare the EED estimation accuracy of SCORE and ROPE in the setting of full-pixel motion compensated coding. A standard H.264 encoder constrained to work only with full-pixel motion compensation resolution generates the bit-stream, while incorporating some error-resilience via random intra coding, i.e., in each frame 10% of the macroblocks are randomly selected to be intra-coded. In parallel, the encoder maintains running distortion estimates via both ROPE and SCORE, solely to evaluate their accuracy. We emphasize that the encoder does not employ these distortion estimates for any optimization of encoding decisions. The transmission of this video sequence is simulated over 100 different realizations of a lossy channel. The distortion of each frame in the video sequence is averaged over realizations. In the case of ROPE, the per-pixel EED estimate is averaged across pixels in a frame, whereas in the case of SCORE the average is across DCT coefficients within a frame. Fig. 2 compares the PSNRs obtained by simulation, and the estimates obtained via SCORE and ROPE. The SCORE estimate matches that of ROPE, and both are very close to the simulation result.

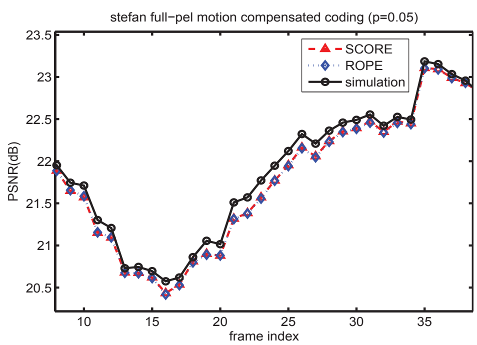

Fig. 2. Comparison of simulated and estimated PSNRs for the $stefan\_cif$ sequence encoded with full-pel motion compensation: bit-rate is $200kbps$, frame rate $30f/s$, and PLR $p = 5\%$.

Next, we compare SCORE and ROPE in the setting of sub-pixel motion compensated coding. The encoder now works at sub-pixel accuracy, and generates distortion estimates via the two modified versions of ROPE that incorporate the Cauchy-Schwarz approximation (Sec.2 and [3]) and pixel distance model (Sec.2 and [5]), respectively, and via SCORE modified in accordance with the approximations in Sec.4. Packet losses are simulated similarly as before. As evident from Fig. 3 SCORE-base technique provides a much more accurate EED estimate than ROPE with Cauchy-Schwarz or pixel distance model approximations. Similar estimation accuracy trends were observed with other video sequences as well, with diverse motion levels.

6. CONCLUSIONS

Our recently introduced technique, SCORE, performs its recursion entirely in transform domain, to find the to estimate end-to-end distortion. SCORE is further extended to accurately account for sub-pixel motion compensation. The scheme exploits the decorrelation property of the transform as well as its energy compaction to closely track the cross correlation introduced by sub-pixel linear interpolation. The efficacy of the new approach is demonstrated via experiments, which indicate that it substantially outperforms the competing pixel domain ROPE variants.

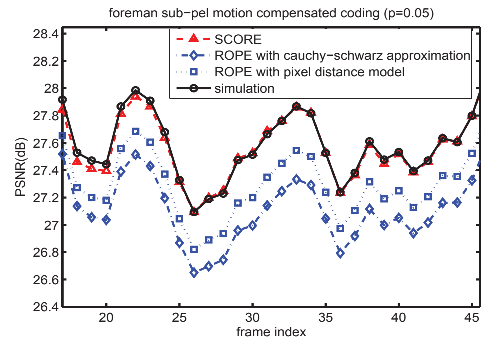

Fig. 3. Comparison of simulated and estimated PSNRs for the $foreman\_cif$ sequence encoded with sub-pel motion compensation: bit-rate is $200kbps$, frame rate $30f/s$, and PLR $p = 5\%$.

7. REFERENCES

[1] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Sig. Proc. Mag.*, vol. 17, no. 4, pp. 61–82, Jul 2000.

[2] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Jrnl. Sel. Areas Comm.*, vol. 18, pp. 966–976, June 2000.

[3] A. Leontaris and P. C. Cosman, "Video compression for lossy packet networks with mode switching and a dual-frame buffer," *IEEE Trans. Img. Proc.*, vol. 13, no. 7, pp. 885897, Jul 2004.

[4] B. A. Heng, J. G. Apostolopoulos, and J. S. Lim, "End-to-end rate-distortion optimized md mode selection for multiple description video coding," in *EURASIP Jrnl. App. Sig. Proc.*, 2006.

[5] H. Yang and K. Rose, "Advances in recursive per-pixel end-to-end distortion estimation for robust video coding in H.264/AVC," *IEEE Trans. Circ. Sys. Video Tech.*, vol. 17, pp. 845–856, July 2007.

[6] T. Wedi and H. G. Musmann, "Motion and aliasing compensated prediction for hybrid video coding," *IEEE Trans. Circ. Sys. Video Tech.*, vol. 13, pp. 577–586, July 2003.

[7] K. Rose and S. L. Regunathan, "Toward optimality in scalable predictive coding," *IEEE Trans. Img. Proc.*, vol. 10, no. 7, pp. 965–976, Jul 2001.

[8] J. Han, V. Melkote, and K. Rose, "Transform-domain temporal prediction in video coding: exploiting correlation variation across coefficients," in *IEEE ICIP*, Sep 2010.

[9] J. Han, V. Melkote, and K. Rose, "Estimation-theoretic delayed decoding of predictively encoded video sequences," in *Proc. IEEE DCC*, Mar 2010.

[10] J. Han, V. Melkote, and K. Rose, "A recursive optimal spectral estimate of end-to-end distortion in video communications," in *to appear in Proc. Packet Video*, Dec 2010.

[11] A. Majumdar, J. Wang, and K. Ramchandran, "Drift reduction in predictive video transmission using a distributed source coded side-channel," in *ACM Multimedia*, Oct 2004.

[12] M. Fumagalli, M. Tagliasacchi, and S. Tubaro, "Improved bit allocation in an error-resilient scheme based on distributed source coding," in *IEEE ICASSP*, May 2006.