BLOCK-SIZE ADAPTIVE TRANSFORM DOMAIN ESTIMATION OF END-TO-END DISTORTION FOR ERROR-RESILIENT VIDEO CODING

Bohan Li, Tejaswi Nanjundaswamy, Kenneth Rose

Department of Electrical and Computer Engineering, University of California Santa Barbara
Email: {bohan_li, tejaswi, rose}@ece.ucsb.edu

ABSTRACT

The accuracy of end-to-end distortion (EED) estimation is crucial to achieving effective error resilient video coding. An established solution, the recursive optimal per-pixel estimate (ROPE), does so by tracking the first and second moments of decoder-reconstructed pixels. An alternative estimation approach, the spectral coefficient-wise optimal recursive estimate (SCORE), tracks instead moments of decoder-reconstructed transform coefficients, which enables accounting for transform domain operations. However, the SCORE formulation relies on a fixed transform block size, which is incompatible with recent standards. This paper proposes a non-trivial generalization of the SCORE framework which, in particular, accounts for arbitrary block size combinations involving the current and reference block partitions. This seemingly intractable objective is achieved by a two-step approach: i) Given the fixed block size moments of a reference frame, estimate moments of transform coefficients for the codec-selected current block partition; ii) Convert the current results to transform coefficient moments corresponding to a regular fixed block size grid, to facilitate EED estimation for the next frame. Experimental results first demonstrate the accuracy of the proposed estimate in conjunction with transform domain temporal prediction. Then the estimate is leveraged to optimize the coding mode and yields considerable gains in rate-distortion performance.

Index Terms— End-to-end distortion, joint source channel coding, variable block size coding

1. INTRODUCTION

In most current video coding systems, motion compensated prediction is employed to exploit temporal redundancies. However, due to the temporal and spatial error propagation via the prediction loop, it also increases the vulnerability to packet loss through channels. To mitigate this problem, many error resilience tools and paradigms have been employed, including forward error correction, intra refresh, multiple description coding, and macro block re-transmission [1]. These error resilience methods typically introduce redundancies in the compressed signal, and hence incur additional bit-rate costs. Therefore, the fundamental optimization problem that underlies the coder is formulated in terms of the trade-off between bit-rate and the distortion experienced at the decoder, also referred to as end-to-end distortion (EED). It is thus obvious that the encoder's ability to accurately estimate the EED, accounting for all factors including compression, packet loss and error propagation, is crucial for the optimization of encoding decisions.

The recursive optimal per-pixel estimate (ROPE) [2], which originated in our lab, is well known as an approach to optimally estimate the EED. The main idea of ROPE is to treat the decoder reconstructed pixels as random variables (due to the randomness of packet losses) and rather than exhaustively simulating the decoding procedure at the encoder [3], ROPE recursively calculates the first and second moments of the reconstructed pixels and then directly obtains the optimal EED estimate. The calculation was done via update equations that explicitly account for motion compensated prediction, packet loss rate, and concealment at the decoder. The basic version of ROPE was extended in [4] to account for operations such as sub-pixel motion compensation, de-blocking, and rounding, where inter-pixel correlation terms are involved. ROPE and its extensions have been successfully incorporated into various error-resilient video coding methods [5, 6, 7, 8].

Since ROPE estimates the EED via pixel-domain calculations, it is inherently restricted to account for error propagation due to recursive operations performed in the pixel domain. However, various source coding approaches of significant interest involve recursive operations in the transform domain. Particularly, estimation-theoretic approaches were proposed in [9, 10, 11, 12], wherein substantial compression gains were achieved by recursively operating in the transform domain, which is typically the discrete cosine transform (DCT) domain. In this paper we are specifically interested in the transform domain temporal prediction (TDTP) scheme, which largely eliminates spatial correlations before spectral components (transform coefficients) are independently predicted. The true temporal correlations, which vary considerably from low to high frequency components, are exploited by the method to offer a better prediction.

To provide a ROPE-like EED estimate capable of accounting for error propagation due to recursive operations in the transform domain, the spectral coefficient-wise optimal recursive estimate (SCORE) was proposed in [13]. In SCORE, the moments of the motion compensated prediction reference block's transform coefficients are estimated using the moments of the previous frame's on-grid blocks' transform coefficients. Since DCT is linear, this estimation is achieved through another linear transformation. Given the moments for the reference block, ROPE-like equations are used to estimate the moments of the current block, which in turn are used to estimate EED. SCORE, with the TDTP scheme, was deployed over a lossy network to show its efficacy. It was also later extended for sub-pixel motion compensation techniques [14].

Note that, since SCORE operates on the transform coefficients, another factor of transform block size is introduced into the mix. A fixed block size was previously used in [13, 14] to good effect in the H.264 video codec [15]. In this setting, the linear transform required to estimate moments for the reference block depends only on the position of the reference block, which limits the number of such transforms. Hence they were all calculated offline and stored, as calculating them on the fly will add significant complexity to the en-

This work was supported in part by LG Electronics Inc.

coding. However, the performance of recent standards, such as High Efficiency Video Coding (HEVC) [16], and VP9 [17], relies heavily on variable block-size schemes, wherein prediction and transform block sizes are adapted based on the content. This variable block size setting poses a significant challenge to implement SCORE. Since different transform block sizes could be employed along a motion trajectory, the linear transform required to estimate moments for the reference block, additionally depends on the size of the reference block and the combination of the transform block sizes it spans. This dramatically increases the number of possible transforms and storing all such transforms is impractical. Thus we propose to overcome this challenge, by generalizing the SCORE framework to account for arbitrary block size combinations. We achieve this by breaking the estimation of moments for the reference block into multiple steps, wherein the moments of previous frame's blocks are first converted to moments of coefficients corresponding to a regular grid of fixed transform block size, and then these moments are converted to those of the reference block. The smallest transform block size is used as the anchor in between, as it limits the number of possible transforms required in any of the steps. Experimental results show that the proposed approach achieves accurate estimation of EED. It is also demonstrated that the accurate estimate of EED for the variable block size setting, translates into rate-distortion (R-D) performance improvements for video transmission over a lossy network.

2. RELEVANT BACKGROUND

2.1. Transform-domain temporal prediction (TDTP)

This section briefly introduces TDTP, as it is employed in this paper as a specific example of transform domain based operation. More details regarding TDTP can be found in [10, 11]. Conventional motion-compensated prediction assumes that a sequence of pixels (from consecutive frames) along a motion trajectory form a temporal AR process, and that such sequences are independent of each other. This assumption clearly ignores the inter-pixel (spatial) correlation. Instead, the TDTP approach models a pair of transform coefficients at a given frequency, denoted by (x_n, x_{n-1}), of an inter-coded block and its motion compensated reference, as two successive samples of a scalar AR process, i.e., $x_n = \rho x_{n-1} + z_n$, where ρ is the correlation coefficient corresponding to that frequency, and z_n is the innovation sequence. Thus in the encoder, the motion compensated reference is transformed and weighted by the frequency dependent correlation coefficients, then the transform domain prediction error is calculated, which is finally quantized and entropy coded. Note that since TDTP estimates the prediction in the transform domain, ROPE is not suitable for EED estimation.

2.2. Fixed-block-size spectral coefficient-wise optimal recursive estimate (SCORE)

Transform domain based approaches, such as TDTP, inspired SCORE [13]. Let the uncoded value of transform coefficient m in block k of frame n be denoted as, $x_n^{k,m}$, and the encoder and decoder reconstructions of the coefficient as, $\bar{x}_n^{k,m}$, and $\hat{x}_n^{k,m}$, respectively. Let $u_n^{k,m}$ denote the uncoded value of coefficient m in this reference block[1]. Note that this reference block is possibly off-grid. The decoder reconstruction of the coefficient is denoted as $\hat{u}_n^{k,m}$. In a lossy channel, similar to ROPE, the encoder considers

[1]While $u_n^{k,m}$ is indexed by n and k to indicate the location in the current frame n, it is in fact a function of pixels in frame $n-1$

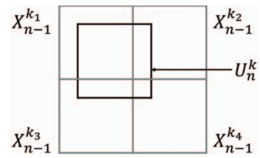


Fig. 1. An example of a off-grid motion compensated block in the fixed block size setting.

$\hat{x}_n^{k,m}$ and $\hat{u}_n^{k,m}$ as random variables. Let the packet loss rate (PLR) be p. The expected distortion at coefficient $x_n^{k,m}$ is

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

which clearly requires the first and second moments of the decoder reconstruction $\hat{x}_n^{k,m}$. SCORE employs two separately developed recursion functions for the cases of intra and inter coding, to compute the moments of each transform coefficients in a frame sequentially. In this paper, we employ the simple "slice copy" error concealment method, wherein if the packet containing the current slice is lost, the co-located reconstruction in the previous frame is copied as the reconstruction of the current slice.

Intra-coding: The recursions are the same as in ROPE, but in transform domain.

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

Inter-coding: Let $\hat{r}_n^{k,m}$ denote the quantized transform coefficient residual. It can be shown that,

$$\begin{aligned} E\{\hat{x}_n^{k,m}\}(P) &= (1-p)(\hat{r}_n^{k,m} + \rho E\{\hat{u}_n^{k,m}\}) + pE\{\hat{x}_{n-1}^{k,m}\}, \\ E\{(\hat{x}_n^{k,m})^2\}(P) &= (1-p)((\hat{r}_n^{k,m})^2 + 2\rho\hat{r}_n^{k,m} E\{\hat{u}_n^{k,m}\} \\ &\quad + \rho^2 E\{(\hat{u}_n^{k,m})^2\}) + pE\{(\hat{x}_{n-1}^{k,m})^2\}, \end{aligned} \tag{2}$$

where ρ is the correlation coefficient in TDTP. As shown in (2), employing TDTP in SCORE is very simple and straight-forward, while basic ROPE cannot account for this transform domain method. Note that, $E\{\hat{u}_n^{k,m}\}$ and $E\{(\hat{u}_n^{k,m})^2\}$ may not be immediately available from the reference frame, since the motion compensated block could be potentially off-grid. Thus these moments need to be calculated from the first and second moments in the reference frame.

If the block size is fixed, any off-grid block in a frame overlaps with at most four on-grid blocks (as illustrated in Fig. 1). Let block U_n^k shown in the figure denote the reference block for the current block k in frame n, and it overlaps with on-grid blocks $X_{n-1}^{k_i}$ in frame $n-1$. Since DCT is a linear transformation, there exist constants $a_m^{i,l}$, named *construction constants*, such that, decoder reconstruction of block U_n^k can be calculated as,

$$\hat{u}_n^{k,m} = \sum_{i=1}^{4} \sum_{l} a_m^{i,l} \hat{x}_{n-1}^{k_i,l}. \tag{3}$$

These constants only depend on the position of U_n^k relative to the on-grid blocks. Given (3), the first and second moments of $\hat{u}_n^{k,m}$ are:

$$\begin{aligned} E\{\hat{u}_n^{k,m}\} &= \sum_{i=1}^{4} \sum_{l} a_m^{i,l} E\{\hat{x}_{n-1}^{k_i,l}\}; \\ E\{(\hat{u}_n^{k,m})^2\} &= \sum_{i=1}^{4} \sum_{i'=1}^{4} \sum_{l} \sum_{l'} a_m^{i,l} a_m^{i',l'} E\{\hat{x}_{n-1}^{k_i,l} \hat{x}_{n-1}^{k_{i'},l'}\}. \end{aligned} \tag{4}$$

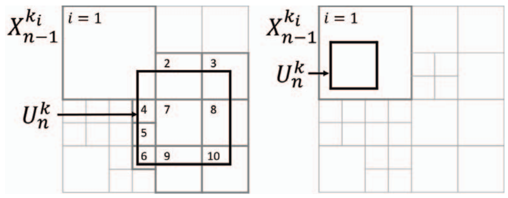


Fig. 2. Examples of off-grid motion compensated blocks in the variable block size setting.

Although cross-correlations are required, the advantage of operating in transform domain is that it largely decorrelates the block. Specifically, as mentioned in [13], the following assumption of 'uncorrelatedness' holds well in the DCT domain:

$$E\{\hat{x}_n^{k_i,l}\hat{x}_n^{k_{i'},l'}\} \approx E\{\hat{x}_n^{k_i,l}\}E\{\hat{x}_n^{k_{i'},l'}\}, \tag{5}$$

if $l \neq l'$, or $i \neq i'$. In this paper, we further assume that the cross correlation between DC coefficients, $\hat{x}_n^{k_i,0}$ and $\hat{x}_n^{k_{i'},0}$, is 1.

3. GENERALIZATION OF SCORE TO VARIABLE BLOCK SIZED CODING

In the fixed block size setting of Sec. 2.2, the construction constants required to estimate moments for the reference block depended only on the position of the reference block, which limited the number of such transforms. Hence they could all be calculated offline and stored to avoid addition of significant complexity to calculate them on the fly during encoding. However, in the variable block size setting, different transform block sizes could be employed along a motion trajectory. Example illustrations of such off-grid reference blocks is shown in Fig. 2. Similar to the fixed block size setting, due to the linearity of DCT, a set of construction constants can be obtained for each pattern to calculate the moments of $\hat{u}_n^{k,m}$ via (4), however, with different number of blocks and size of each block. Unfortunately, the construction constants depend not only on the position of the block U_n^k, but depend also on the size of the block U_n^k, and the size and position of the blocks $X_{n-1}^{k_i}$ it spans. This dramatically increases the number of possible off-grid patterns and storing all such construction constants is impractical.

To overcome this challenge, we propose a general approach to account for any arbitrary combination of block sizes and positions, while still leveraging the advantages observed in the fixed block size setting. Specifically, we propose to break the estimation of moments for the reference block into multiple steps (as shown in Fig. 3):

Step 1: We break the blocks $X_{n-1}^{k_i}$ to a regular grid of blocks with the minimum transform block size (which is 4x4 in this paper), and calculate the moments of the new transform coefficients. Let $Y_{n-1}^{k_{ij}}$ and $\hat{y}_{n-1}^{k_{ij},f}$ denote the small block j and its coefficients, respectively. Again due to linearity of DCT, a set of constants $b_{j,f}^l$ exists for each coefficient f, such that,

$$\hat{y}_{n-1}^{k_{ij},f} = \sum_l b_{j,f}^l \hat{x}_{n-1}^{k_i,l}. \tag{6}$$

Thus the first and second moments of $\hat{y}_{n-1}^{k_{ij},f}$ are calculated via equations similar to (4).

Step 2: We perform motion compensation in this fixed block size setting. Given the moments of $\hat{y}_{n-1}^{k_{ij},f}$, we calculate moments of the potentially off-grid coefficients according to Sec. 2.2. The motion compensated blocks and their coefficients are denoted as, $V_n^{k_h}$, and $\hat{v}_n^{k_h,f}$, respectively.

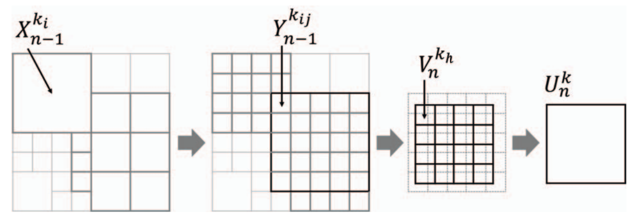


Fig. 3. Illustration of the proposed method. For each off-grid reference block, the blocks it spans are broken into 4x4 blocks, off-grid adjusted, and finally combined to the required block size. Note that each step is an estimation of the first and second moments of the corresponding transform coefficients.

Step 3: We finally combine the small blocks $V_n^{k_h}$ back to the required size, and calculate the moments of the new coefficients. The block we need, U_n^k, and the coefficients $\hat{u}_n^{k,m}$ can be represented in terms of $\hat{v}_n^{k_h,f}$ as:

$$\hat{u}_n^{k,m} = \sum_h \sum_f c_m^{h,f} \hat{v}_n^{k_h,f}, \tag{7}$$

where $c_m^{h,f}$ is another set of construction constants. Using these relations we calculate the moments of $\hat{u}_n^{k,m}$.

Overall we require three sets of construction constants, one set that is same as Sec. 2.2, and two additional sets for step 1, $b_{j,f}^l$, and step 3, $c_m^{h,f}$. However, breaking and combining blocks results in very limited patterns, leading to reduction of complexity and simplification of implementation. Note that similar to Sec. 2.2, we make an assumption of 'uncorrelatedness' in each step for estimation of second moments, which is shown as a valid assumption in the results.

Within the encoder, we perform step 1 right after a frame is encoded, and moments of $\hat{y}_n^{k_{ij},f}$ are buffered for the next frame. That is, the first and second moments are maintained in a fixed block size manner, and step 2 and 3 are employed to adapt them to variable block sizes. Given the moments for $\hat{u}_n^{k,m}$, EED is estimated according to (2).

3.1. Deployment for encoder mode-selection

The EED estimate provided by the proposed method is then employed in the mode-selection (Inter/Intra) optimization at the encoder. For a coding block k in frame n, with coding mode μ and quantization parameter q, we denote the EED and bit cost as, $D_n^k(q,\mu)$ and $B_n^k(q,\mu)$, respectively. The problem of optimizing the coding mode per block, given the quantization parameter, is given as:

$$\mu_n^k(\lambda, q_n) = \arg\min_{\mu}\{D_n^k(q_n,\mu) + \lambda B_n^k(q_n,\mu)\}, \tag{8}$$

where λ is a Lagrange parameter whose value is fixed for all frames in the simulation. Varying the value of λ results in an operational rate-distortion curve. Experimental results demonstrate that employing the proposed SCORE in the variable block size setting, results in better R-D performance in a lossy channel.

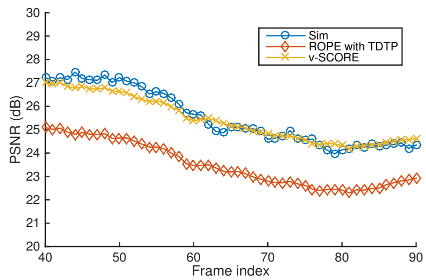


Fig. 4. Comparison of PSNR estimates and actual PSNR averaged over 100 simulated packet loss patterns for the BQTerrace sequence (1920x1080) at the PLR of $p = 5\%$.

4. EXPERIMENTAL RESULTS

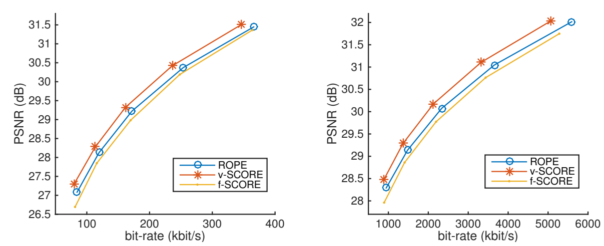
In our experiments, ROPE, fixed-block-size SCORE (f-SCORE) and variable-block-size SCORE (v-SCORE) are implemented in an HEVC encoder. TDTP is employed in SCORE based coders. Slice-copy error concealment is employed in all the coders. In the fixed block size setting, the transform block size is set to 4x4, while in variable block size setting, the transform block size may vary from 4x4 to 16x16. Note that in this paper, we constrain the encoder to full-pixel motion compensation to demonstrate the potential of the proposed approach. Very recent advances in TDTP [11] extended its ability to provide considerable gains to the setting of sub pixel motion, and while the results herein are in conjunction with the original full pixel TDTP to provide the proof of concept for block size adaptive TDTP with efficient EED estimate, ongoing work is underway to combine the approach with sub pixel TDTP.

We first evaluate the EED estimation accuracy of v-SCORE in coders that employ the 'random intra' error-resilience technique, where in each frame, 10% of coding units are randomly selected to be intra-coded. That is, the coders do not employ the distortion estimates for any optimization of encoding decisions. Here, TDTP is employed in both ROPE and v-SCORE coders, and we calculate the distortion in ROPE by averaging the per-pixel EED estimate, and in v-SCORE by averaging the per-coefficient estimate. We also simulate the transmission of the bitstream over 100 different realizations of a lossy channel, and average the distortion over realizations for each frame to obtain a simulation result. The ROPE and SCORE estimates are compared with the simulation result to evaluate their accuracy. The PSNRs obtained by simulation, ROPE, and v-SCORE are shown in Fig. 4. It can be seen that ROPE is not capable of accounting for the transform domain operations of TDTP, hence it significantly underestimates the quality at the decoder. However, v-SCORE yields fairly accurate estimation compared to the simulation result.

Next, we compare the R-D performance of employing ROPE, f-SCORE and v-SCORE for mode-selection optimization as described in Sec. 3.1. The TDTP correlation coefficients employed were estimated via methods presented in [11], from a training set outside the test set for which results are reported. Note that TDTP is *not* employed in ROPE for this experiment. In Fig. 5a and Fig. 5b, the R-D curves at PLR of 5% for the competing coders are shown for two video sequences of different resolutions, CIF and HD. Percentage BD-rate [18] reduction obtained by v-SCORE, over f-SCORE (Exp v-f) and over ROPE (Exp v-R), at PLR of 1% and 5% are presented in Table 1. By comparing the results of v-SCORE and f-SCORE, it can be concluded that significant gains can be achieved by generalizing SCORE to the variable block size setting via the proposed approach. Note the performance gain is more substantial for HD sequences, since unlike low-resolution sequences, where a 4x4 block size itself may be optimal, the block sizes may vary to a large extent for high-resolution sequences. Note that ROPE out-performing f-SCORE also demonstrates the severe limitation of fixed block size coding, where in despite employing TDTP, f-SCORE lags behind ROPE without TDTP but with variable block sizes. Finally, the gains of v-SCORE over ROPE demonstrates the utility of employing transform domain based techniques, specifically, TDTP in this paper. In all, considerable performance gain is achieved for error-resilient video coding. Note that the complexity of v-SCORE is higher than that of ROPE due to the additional step of conversion from variable to fixed block size partition. A fair complexity comparison requires optimization of the v-SCORE code to match the maturity of the ROPE code, a work that is currently under way.

Table 1. Bit-rate reduction (%) of v-SCORE compared to f-SCORE and ROPE

	PLR $p = 1\%$		PLR $p = 5\%$	
Sequence	Exp v-f	Exp v-R	Exp v-f	Exp v-R
coastguard_cif	29.78	11.14	28.04	8.38
bus_cif	15.37	8.48	13.04	4.37
foreman_cif	15.76	8.90	13.94	9.20
flower_cif	10.32	12.56	8.44	8.12
BasketballDrive	32.06	4.46	32.76	4.35
BQTerrace	23.11	20.06	21.80	19.18
Kimono	31.02	5.54	28.49	4.18
ParkScene	19.61	22.92	19.47	12.86
Average	22.13	11.76	20.75	8.83



(a) foreman (CIF) (b) ParkScene (1920x1080)

Fig. 5. R-D curves for HEVC coders with mode decision optimization via v-SCORE, f-SCORE, and ROPE, at PLR $p = 5\%$, for two sequences of different resolutions.

5. CONCLUSION

In this paper, a novel approach is proposed to generalize the EED estimation method, SCORE, to account for the variable block size setting. This approach enables substantial gains from employing variable transform block sizes, while maintaining simplicity and accuracy. Experimental results show that the proposed method provides accurate EED estimation, and that it enables considerable R-D performance improvement for error-resilient video coding.

6. REFERENCES

[1] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Sig. Proc. Mag.*, vol. 17, 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, Jun 2000.

[3] T. Stockhammer, T. Wiegand, and S. Wenger, "Optimized transmission of h. 26l/jvt coded video over packet-lossy networks," in *IEEE ICIP*, 2002.

[4] 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, Jul 2007.

[5] 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, pp. 885–897, Jul 2004.

[6] B. A. Heng, J. G. Apostolopoulos, and J. S. Lim, "End-to-End Rate-Distortion Optimized MD Mode Selection for Multiple Description Video Coding," *EURASIP Jrnl. App. Sig. Proc.*, 2006.

[7] A. R. Reibman, L. Bottou, and A. Basso, "Dct-based scalable video coding with drift," in *IEEE ICIP*, Oct 2001.

[8] F. Zhai, C. E. Luna, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and packet classification for real-time video transmission over differentiated services networks," *IEEE Trans. Multimedia*, vol. 7, pp. 716–726, Aug 2005.

[9] K. Rose and S. L. Regunathan, "Toward optimality in scalable predictive coding," *IEEE Trans. Img. Proc.*, vol. 10, pp. 965–976, Jul 2001.

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

[11] S. Li, T. Nanjundaswamy, Y. Chen, and K. Rose, "Asymptotic closed-loop design for transform domain temporal prediction," in *IEEE ICIP*, Sep 2015.

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

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

[14] J. Han, V. Melkote, and K. Rose, "A spectral approach to recursive end-to-end distortion estimation for sub-pixel motion-compensated video coding," in *IEEE ICASSP*, May 2011.

[15] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the h. 264/avc video coding standard," *IEEE Trans. Circ. Sys. Video Tech.*, vol. 13, pp. 560–576, Jul 2003.

[16] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Trans. Circ. Sys. Video Tech.*, vol. 22, pp. 1649–1668, Dec 2012.

[17] J. Bankoski, R. S. Bultje, A. Grange, Q. Gu, J. Han, J. Koleszar, D. Mukherjee, P. Wilkins, and Y. Xu, "Towards a next generation open-source video codec," in *Proc. SPIE 8666, Visual Information Processing and Communication IV, 866606*, Feb 2013.

[18] G. Bjontegaard, "Calcuation of average psnr differences between rd-curves," *Doc. VCEG-M33 ITU-T Q6/16, Austin, TX, USA, 2-4*, Apr 2001.