AN ERROR-RESILIENT VIDEO CODING FRAMEWORK WITH SOFT RESET AND END-TO-END DISTORTION OPTIMIZATION

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

Temporal prediction plays a crucial role in most video coding applications. However, due to error propagation via the prediction loop, it also increases the vulnerability to channel loss. The standard counter measure to mitigate error propagation is the 'intra refresh' mode, which in effect resets temporal prediction to block error propagation, but at a significant rate overhead. This on/off switch for temporal prediction is overly crude to optimize the compression-resilience tradeoff. In this paper, we propose a novel framework that significantly expands the options available to counter error propagation by introducing optimally controlled soft resets, wherein intra and inter predictions are combined with adjustable weights to control the dependency on previous frames while accounting for the overall rate and distortion. Since the optimal control of such soft resets can only be achieved if the encoder can effectively estimate its impact on the end-to-end distortion (EED), we propose to extend the well known recursive optimal per-pixel estimation (ROPE) approach to accurately account for the soft reset mode, then optimize encoder mode decisions to minimize the estimated EED for the given rate. Experimental results show that the proposed framework achieves significant performance gains for video streaming over unreliable networks.

Index Terms— End-to-end distortion, joint source channel coding, soft reset

1. INTRODUCTION

In most current video coding systems, motion compensated prediction is employed to exploit temporal redundancies by effectively predicting the content of the current frame from previous reconstructed frames. However, the extensive benefits of temporal prediction come at considerable cost when the encoded data are transmitted over unreliable networks, as errors introduced due to packet loss propagate through the prediction loop and impact future frames, causing a substantial and extended degradation of quality. 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]. Specifically, we focus on the intra refresh method (also referred to as the constrained intra prediction) among all error resilience methods above, since it is the only one addressing the error propagation directly, while others attempt to either reduce the effective rate of packet loss or mitigate its immediate impact.

In contrast to the inter prediction mode, the intra refresh mode predicts only from information available in the current packet, resulting in an instant reset of error propagation. However, as typical error resilience methods usually do, it also introduces redundancies in the compressed signal, and hence incurs additional bit-rate costs. Therefore, when deciding between the inter and intra refresh modes, 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 in order to optimize the encoding decisions, it is crucial for the encoder to accurately estimate the EED, accounting for all factors including compression, packet loss and error propagation.

Our lab's recursive optimal per-pixel estimation (ROPE) [2] approach is well known 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 is 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].

With accurately estimated EED by ROPE, the encoder can optimally switch between the inter prediction and the intra refresh modes to account for the trade off between compression and error resilience. However, although mode selection between the above two modes with EED estimation provides good performance improvement [2], the 'intra refresh' approach is in fact the outcome of repurposing the existing tool of intra coding, originally designed to encode newly appearing objects which are not predictable from the past, to arbitrarily reset temporal prediction (and hence stop error propagation) at a significant cost in rate overhead. Such ad-hoc repurposing of existing tools is significantly suboptimal.

In this paper, a novel framework is proposed, where options to control the error propagation are significantly expanded, including a soft reset joint intra-inter prediction mode specifically designed for a controlled tradeoff between compression and resilience. As a crucial part of the framework, accurate EED estimation methods for the expanded options are also developed, in order to enable the encoder to optimally decide its mode selection and thus take advantage of the broader options of controlling error propagation. Experimental results show that compared to mode selection among existing prediction modes, the proposed framework with soft reset prediction mode yields a significant gain for video streaming over lossy channels.

This work was supported in part by Google Inc.

2. RELEVANT BACKGROUND

Consider point-to-point video communication, assuming that packet loss is statistically uniformly distributed with packet loss rate p available to the encoder (for simplicity but without loss of generality, since extensions of ROPE have been developed for different network models [6, 9] and can be generalized to the EED estimation methods introduced in this paper). For optimal performance, the encoder must optimize its decisions with respect to the *decoder* reconstructed video quality. However, the decoder reconstruction is a random process as far as the encoder is concerned, with the influence of channel loss greatly complicated by error propagation through the prediction loop, error concealment efforts at the decoder, etc.

Therefore, ROPE considers the decoder reconstruction of each pixel as a random variable, and estimates the expected EED. Let the uncoded value of the pixel at location m in block k of frame n be denoted as $x^m_{n,k}$, and the decoder reconstruction of the pixel as $\hat{x}^m_{n,k}$. With mean squared error (MSE) distortion, the expected EED of $x^m_{n,k}$ can be formed as:

$$E\{(x^m_{n,k} - \hat{x}^m_{n,k})^2\} = (x^m_{n,k})^2 + 2x^m_{n,k}E\{\hat{x}^m_{n,k}\} + E\{(\hat{x}^m_{n,k})^2\},$$

which clearly only requires the first and second moments of the decoder reconstruction $\hat{x}^m_{n,k}$.

Thus in order to accurately estimate EED, ROPE recursively tracks the moments of the decoder reconstruction. Note that different decoder error concealment methods and encoder schemes may result in different ROPE recursion formulas. In this paper, we employ the simple 'slice copy' error concealment method, where 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. We further assume each packet contains one frame for simplicity.

For the 'intra refresh' mode, the current block is only predicted using other pixels encoded by the intra refresh mode in the same packet. Thus, as long as the packet containing the current frame is correctly received, the decoder will be able to reconstruct the current block exactly as the encoder reconstruction (denoted as $\bar{x}^m_{n,k}$). Therefore, the recursion formula to track the first and second moments are:

$$\begin{aligned} E\{\hat{x}^m_{n,k}\} &= (1-p)\bar{x}^m_{n,k} + pE\{\hat{x}^m_{n-1,k}\}, \\ E\{(\hat{x}^m_{n,k})^2\} &= (1-p)(\bar{x}^m_{n,k})^2 + pE\{(\hat{x}^m_{n-1,k})^2\} \end{aligned} \quad (1)$$

For the inter prediction mode, the current block k is predicted by the decoder reconstruction of another block k' in the previous frame (first-order inter prediction is assumed in this paper without loss of generality). Denoting the quantized residual as $\hat{r}^m_{n,k}$, it can be shown that the ROPE recursions for inter prediction are:

$$\begin{aligned} E\{\hat{x}^m_{n,k}\} &= (1-p)(E\{\hat{x}^m_{n-1,k'}\} + \hat{r}^m_{n,k}) + pE\{\hat{x}^m_{n-1,k}\}, \\ E\{(\hat{x}^m_{n,k})^2\} &= (1-p)(E\{(\hat{x}^m_{n-1,k'})^2\} + 2\hat{r}^m_{n,k}E\{\hat{x}^m_{n-1,k'}\} \\ &\quad + (\hat{r}^m_{n,k})^2) + pE\{(\hat{x}^m_{n-1,k})^2\} \end{aligned} \quad (2)$$

As shown in (1) and (2), the first and second moments of decoder reconstructed pixels in the current frame depends only on the moments of decoder reconstructions in the previous frame, thus establishing a recursive method to track the moments, and therefore accurately estimate the EED.

Note that in this paper, only integer precision motion compensation is considered for the purpose of simple presentation. Accurate EED estimation of sub-pixel inter prediction methods have been developed in [4].

3. PROPOSED FRAMEWORK WITH SOFT RESET

As explained in Section 2, with the ability to accurately estimate EED, the encoder is capable of optimally switching between the inter prediction mode, which causes error propagation through the temporal prediction loop, and the intra refresh mode, which fully stops the propagation at that instant. These two modes in effect serve merely as an on/off switch for error propagation, providing a very crude control to the encoder, when in fact with accurate estimate of EED in hand, the encoder can optimally control the extent of error propagation.

Therefore, we propose an error-resilient video coding framework, where besides the inter and intra refresh modes, more options of controlling the error propagation are allowed, and the encoder decisions are based on the EED estimation, thus providing a more flexible control over the trade-off between error-resilience and compression. Specifically, in addition to the inter mode and the intra refresh mode, the *unconstrained intra prediction* mode is first included to provide the option of allowing error propagation through the spatial prediction loop. More importantly, we propose to include the *soft-reset joint inter-intra prediction* mode in order to provide a finer control over error propagation. In the rest of this section, the above two modes and their corresponding methods to overcome the challenges of accurately estimating EED are presented.

3.1. Unconstrained Intra Prediction

While the constrained intra prediction (intra refresh) mode is widely used by error-resilient video coding applications, the unconstrained intra prediction mode, wherein the current block is allowed to be predicted from previously reconstructed inter-predicted pixels within the same frame, is usually the default intra prediction method in non error-resilient coders. The unconstrained intra mode is more efficient in exploiting spatial correlations between blocks, but unlike the constrained intra mode, it suffers from error propagation through the spatial prediction loop, i.e., errors in spatial neighbors of the current block (potentially through temporal error propagation) will influence its intra prediction. We introduce the unconstrained intra mode into our proposed framework as an optional mode to let the encoder have control over the possible error propagation paths. Moreover, as to be seen in Section 3.2, it is also part of our soft reset joint prediction mode.

The EED estimation of unconstrained intra prediction is obviously different from that of constrained intra prediction shown in (1). If the packet containing the current frame is received, the unconstrained intra prediction $\tilde{x}^m_{n,k}(I)$ is a filtered output of previous decoder reconstructions of its neighboring blocks:

$$\tilde{x}^m_{n,k}(I) = \sum_i a_i \hat{x}^{m_i}_{n,k_i}(r), \quad (3)$$

where a_i are the filter coefficients. The decoder reconstruction of the ith reference, given the current frame is correctly received, is denoted as $\hat{x}^{m_i}_{n,k_i}(r)$. If e.g., this sample was reconstructed via inter prediction mode in (2), $\hat{x}^{m_i}_{n,k_i}(r) = \hat{x}^{m_i}_{n-1,k'_i} + \hat{r}^{m_i}_{n,k_i}$.

For the unconstrained intra mode, the moment estimation recursions can then be expressed as:

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

Note that substituting (3) into the second moment $E\{(\tilde{x}_{n,k}^m(I))^2\}$ shows requirement of the cross correlation term $E\{\hat{x}_{n,k_i}^{m_i}(r)\hat{x}_{n,k_j}^{m_j}(r)\}$. Since only the first and second marginal moments are available, we need to approximate the spatial correlation coeffcient ρ_s. In [4], the 'exponential decay' correlation model is presented for EED estimation:

$$\rho_s(d) \approx \exp(-\alpha d), \tag{5}$$

where α is a parameter whose typical value is around 0.05, and d is the distance between the pixels. With this model utilized, we can now estimate the first and second moments of pixels predicted by the unconstrained intra mode, and thus estimate EED accordingly.

3.2. Soft Reset Joint Inter-Intra Prediction

As discussed in Section 3.1, the unconstrained intra prediction mode provides the encoder with an alternate error propagation path, but still does not provide the encoder with a fine control over the degree of error propagation.

To provide a controllable 'soft reset' for the error propagation, we propose to utilize the weighted average of unconstrained intra prediction and inter prediction, namely the joint inter-intra prediction. The joint prediction of pixel $x_{n,k}^m$ can be expressed as:

$$\tilde{x}_{n,k}^m = w^m(P)\tilde{x}_{n,k}^m(P) + w^m(I)\tilde{x}_{n,k}^m(I), \tag{6}$$

where $\tilde{x}_{n,k}^m(P)$ is the inter prediction and $\tilde{x}_{n,k}^m(I)$ is the unconstrained intra prediction, and $w^m(P)$, $w^m(I)$ are the weights for the two predictions, respectively.

Although similar joint inter-intra prediction methods (also referred to as combined inter-intra prediction) have been previously proposed [10, 11], we emphasize here that we are proposing to perform joint inter-intra prediction with both a different motivation and different optimization approach.

First, our goal of using the joint prediction is not for a better prediction. Rather, we average intra and inter prediction in order to provide a soft reset. Recognizing the difference in motivation, we refer to our proposed prediction mode as the 'soft reset joint prediction mode'. Also, due to the different motivations, the weights should not be targeted to address the correlation between the referenced and the predicted pixels, but should be designed for the balance between error resilience and coding efficiency.

Furthermore, since our soft-reset joint prediction is intended for video coding over a lossy channel, establishing a ROPE-like EED estimation method for the joint prediction is crucial for optimal rate-distortion (R-D) decisions in our framework. However, extending ROPE to account for the proposed mode is not trivial. To overcome the challenges, we propose the following methods to estimate EED accurately for the soft-reset joint inter-intra prediction mode.

Similar to (4), we first establish the moment estimation recursions as following:

$$\begin{aligned} E\{\hat{x}_{n,k}^m\} &= (1-p)(E\{\tilde{x}_{n,k}^m\} + \hat{r}_{n,k}^m) + pE\{\hat{x}_{n-1,k}^m\}, \\ E\{(\hat{x}_{n,k}^m)^2\} &= (1-p)(E\{(\tilde{x}_{n,k}^m)^2\} + 2\hat{r}_{n,k}^m E\{\tilde{x}_{n,k}^m\} \\ &\quad + (\hat{r}_{n,k}^m)^2) + pE\{(\hat{x}_{n-1,k}^m)^2\}. \end{aligned} \tag{7}$$

It is obvious that the first and second moments of the joint prediction are needed. Substuting (6) into the moments we have:

$$\begin{aligned} E\{\tilde{x}_{n,k}^m\} =& w^m(P)E\{\tilde{x}_{n,k}^m(P)\} + w^m(I)E\{\tilde{x}_{n,k}^m(I)\}, \\ E\{(\tilde{x}_{n,k}^m)^2\} =& (w^m(P))^2 E\{(\tilde{x}_{n,k}^m(P))^2\} \\ &+ (w^m(I))^2 E\{(\tilde{x}_{n,k}^m(I))^2\} \\ &+ 2w^m(P)w^m(I)E\{\tilde{x}_{n,k}^m(P)\tilde{x}_{n,k}^m(I)\}, \end{aligned} \tag{8}$$

where the first and second moments of $\tilde{x}_{n,k}^m(I)$ are given in Section 3.1 and moments of $\tilde{x}_{n,k}^m(P)$ are given by moments of pixels in the previous frame.

However, note that the correlation term is still not directly available to the encoder. Moreover, the inter prediction is a reconstructed pixel in the previous frame along the motion trajectory, while the intra prediction is a reconstruction in the current frame located at the boundaries of the current block. Therefore, unlike the correlation term in Section 3.1, this correlation of inter and intra prediction, $E\{\tilde{x}_{n,k}^m(P)\tilde{x}_{n,k}^m(I)\}$, is actually a combination of spatial correlation and temporal correlation.

In order to approximate the correlation term with both satisfactory accuracy and complexity, we make the 'separate correlation' assumption, wherein the correlation coefficient ρ is given by the product of the temporal correlation coefficient ρ_t and the spatial correlation coefficient ρ_s:

$$\rho(t, d) \approx \rho_t(t)\rho_s(d). \tag{9}$$

For the temporal correlation, the Markov model can be applied to pixels along the motion trajectory. Since first order temporal prediction is assumed, the time difference t is a constant, thus in this paper we apply ρ_t as a constant (typically 0.95-0.98). The spatial correlation coefficient can be calculated through (5), where the distance d is defined as the distance from the boundary to the predicted pixel along the predicting direction for angular prediction, and the average distance to the upper and left boundary for DC and planar prediction. As will be shown in Section 4, with properly set parameters, this simple approximation of cross correlation is sufficient for accurate EED estimation. With the correlation term estimated, we can finally estimate the moments in (8), and thus estimate the EED for the soft reset joint prediction mode.

Note that for the soft reset joint prediction mode, the unconstrained intra is used rather than intra refresh. This decision is based on the following considerations. First, if the boundaries or even a portion of them are reliable (e.g., coded by the intra refresh mode), the unconstrained intra prediction portion will reasonably reset the error propagation through spatial prediction loop. Second, even when the boundaries are not reliable, the spatial error propagation path usually provides better error-resilience, which ensures the effectiveness of the joint prediction mode to serve as a soft reset. Finally, using intra refresh in joint prediction would severely constrain its ability since it would have significant bit-rate overhead due to the very limited availability of boundaries with no impact of error propagation.

It should also be noted that the purpose of adding the unconstrained intra mode and the soft reset joint prediction mode is to demonstrate the potential of our proposed error resilient video coding framework with EED estimation. The framework is general and effective even if other possible methods to control error propagation are introduced with proper EED estimation.

4. RESULTS AND DISCUSSION

In our experiments, the proposed framework and methods of EED estimation were implemented in the High Efficiency Video Coding (HEVC) [12] reference software and used for the R-D optimization of mode selection. A wide range of video sequences with resolutions ranging from 240P to 1080P were tested. For each video sequence, the first 100 frames were encoded with QP values of 27, 32, 37 and 42. The channel loss was simulated with 100 realizations at a packet loss rate of 5%. The decoder was implemented with the simple 'slice copy' error concealment method and the video coding performance was assessed by averaging the MSE of the decoder reconstructions over the 100 realizations for each sequence and each QP.

To show the performance of our proposed error-resilient video coding framework, three sets of experiments were conducted with different availability of modes. First, the *baseline* includes only the inter and intra refresh modes. In the second set of experiments (denoted as *base+UI*), the unconstrained intra mode is enabled in addition to the two modes in baseline. Finally, in the third set (denoted as *base+UI+soft*), the proposed soft reset joint prediction mode is enabled along with the inter, intra refresh and unconstrained intra modes. For the soft reset mode in the third set, the weights of inter and intra prediction are both set to be 0.5, in order to provide a soft reset where both the predictions have the same importance.

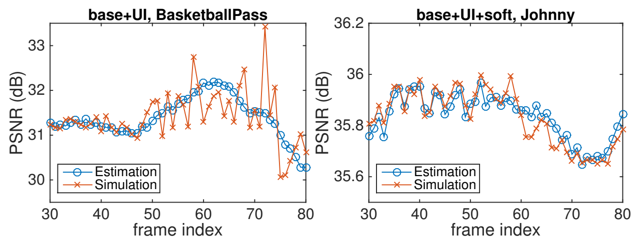

Fig. 1. EED estimation compared with simulated ground truth.

We first show results to compare the estimated EED of each frame with the simulated EED (which can be viewed as the ground truth) to illustrate the accuracy of our proposed EED estimation methods. As shown in Fig. 1, EED estimation of both the base+UI and base+UI+soft settings is quite accurate and follows the general trend seen in simulated results, which confirms that the various assumptions and models introduced in Section 3 are valid for our purpose.

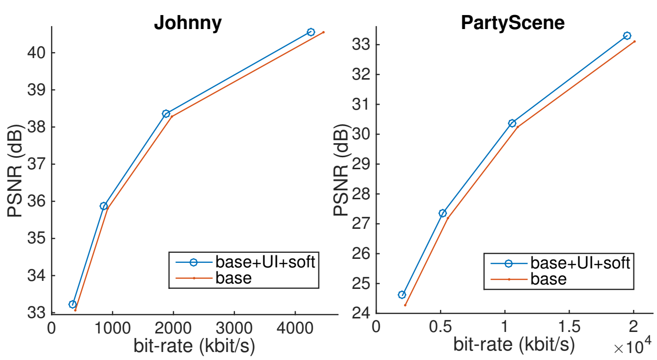

Fig. 2. R-D curves of the proposed base+UI+soft set compared to the baseline.

Next, to demonstrate the performance of our proposed framework, compared to the baseline, the BD-rate [13] reduction of the base+UI set and base+UI+soft set are shown in Table 1. As seen in the table, with unconstrained intra mode enabled we achieve an average BD-rate reduction of 1.72%. On the one hand, this performance gain shows the potential of our framework which benefits from the ability to switch between error propagation paths while accounting for the channel loss with EED estimation. On the other hand, the relatively small gain also illustrates the need for a better designed error-resilient prediction mode.

Table 1. Bit-rate reduction (%) compared to baseline

Sequence	base+UI	base+UI+soft
mobile (CIF)	1.52	9.61
foreman (CIF)	1.75	7.00
flower (CIF)	0.03	3.71
BasketballPass (416×240)	2.55	4.66
BlowingBubbles (416×240)	2.45	12.83
PartyScene (832×480)	0.54	10.04
FourPeople (1280×720)	2.07	5.45
Johnny (1280×720)	5.90	8.76
BQTerrace (1920×1080)	0.03	2.76
ParkScene (1920×1080)	0.36	6.13
Average	1.72	7.09

This need is confirmed by the results of the base+UI+soft set, which achieves a significant average **BD-rate reduction of 7.09%** due to the introduction of the proposed soft reset joint prediction mode. The R-D curves comparing the base+UI+soft set to the baseline of two sequences are also shown in Fig. 2, which confirm its effectiveness for a wide range of operating points. Overall, the results show that with properly designed options, our framework provides considerable performance gain for video streaming over lossy networks.

Note in our experiments, the parameter α in (5), as well the value of temporal correlation coefficient ρ_t are manually selected for each sequence and are set as constants for the test 100 frames. This is clearly suboptimal since it not only requires manual adjustment, but also ignores the fact that video content statistics are not guaranteed to be stationary, either within a single frame or across multiple frames. To address this problem, on-going research is focused on estimating even these parameters recursively for every pixel, which allows the encoder to capture the local statistics.

It should also be noted that, in our experiments, the weights in the soft reset joint prediction mode are chosen as a constant value of 0.5. Although the current results already show a significant performance gain, the weights should be better designed to accommodate different video content, block size, bit-rate, packet loss rate, etc. Hence, on another front, on-going work is focused on various approaches of designing the weights, which could be introduced as multiple options of weight combinations in the proposed framework.

5. CONCLUSION

In this paper, a novel error-resilient video coding framework with EED estimation and soft reset joint prediction is proposed. With the framework, the encoder's options to finely control the error propagation are significantly expanded. Furthermore, with accurate EED estimation for each option, the encoder is capable of selecting the modes while accounting for both the overall bit-rate and error resilience. Experimental results show that the proposed framework with soft reset joint prediction achieves considerable performance gains for video streaming over lossy networks.

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 voding," *EURASIP Jrnl. App. Sig. Proc.*, vol. 2006, pp. 261–261, 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] Y. Liao and J. D. Gibson, "Rate-distortion based mode selection for video coding over wireless networkswith burst losses," in *17th International Packet Video Workshop*, May 2009.

[10] J. Xin, K. N. Ngan, and G. Zhu, "Combined inter-intra prediction for high definition video coding," in *Picture Coding Symposium*, 2007.

[11] Y. Chen, K. Rose, J. Han, and D. Mukherjee, "A pre-filtering approach to exploit decoupled prediction and transform block structures in video coding," in *IEEE ICIP*, Oct 2014.

[12] 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.

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