Optimal mode switching for multi-hypothesis motion compensated prediction

Ramdas Satyan #1, Fabrice Labeau #2, Kenneth Rose *3

Electrical and Computer Engineering, McGill University
Montreal, Quebec H3A 2A7,Canada
[1] ramdas.satyan@mail.mcgill.ca
[2] fabrice.labeau@mcgill.ca

* *Electrical and Computer Engineering, University of California Santa Barbara*
Santa Barbara, CA 93106, USA
[3] rose@ece.ucsb.edu

***Abstract*—Transmission of compressed video over unreliable networks is vulnerable to errors and error propagation. Multi-hypothesis motion compensated prediction (MHMCP) which was originally developed to improve compression efficiency has been shown to have a good error resilience property. In this paper we improve the overall performance of MHMCP in packet loss scenarios by performing optimal mode switching within a rate distortion framework. The approach builds on the recursive optimal per-pixel estimate (ROPE), which is extended by re-deriving recursion formulas for the more complex MHMCP setting, so as to achieve an accurate estimation of the end-to-end distortion. Simulation results show significant performance gains over the standard MHMCP scheme and the importance of performing effective mode decisions. We also show results with comparison to conventional ROPE.**

I. Introduction

In recent years video communication over packet switched networks such as Internet has become universal and overwhelmingly important. These networks currently provide very limited or no end-to-end Quality of Service. Transmission of compressed video over these networks is thus highly susceptible to errors and packet losses. The use of motion compensated prediction in video coding causes these errors to propagate to subsequent frames. This problem is severe and leads to a substantial deterioration of received video quality. Considerable research has taken place to make video communication more robust to these conditions.

The most common and effective way of mitigating the effect of packet loss is to switch off the inter-frame prediction loop for certain macroblocks (MBs). This is refereed to as Intra coding and these MBs no longer depend on the past frames and error propagation is stopped. An Intra MB generally requires higher bit-rate than an Inter MB and too many Intra MBs will drastically reduce the compression efficiency. To achieve the right balance between coding efficiency and error resilience switching between intra and inter coding is very critical. A complete solution to this problem for a single reference frame [1] within an overall rate-distortion (RD) framework has been proposed. A mode switching algorithm with two reference frames [2] (one short term and one long term frame) has been proposed to improve the robustness of the compressed bitstream. Each MB is predicted using one of these reference frames. Both techniques [1], [2] use ROPE to estimate the overall end-to-end distortion and a single hypothesis to predict each MB.

When a linear combination of multiple signals (hypotheses) is used to predict a MB then it is termed as MHMCP. Each hypothesis can be from the same reference frame or multiple reference frames. MHMCP was originally developed to improve the compression efficiency of video coding over single hypothesis [3], [4]. It has an inherent resilience property in error prone environments where error propagation is reduced by performing prediction from several hypothesis. When a hypothesis is corrupted during transmission then other uncorrupted hypotheses can be used to make a reliable prediction and thereby reducing error propagation. Its error resilience capability was analyzed in [5], where a two-hypothesis MCP is used. In this approach a frame level model for the decoder distortion due to error propagation has been developed and combined with the encoder predictor to find the trade-off between coding efficiency and error resilience. An extension to the two-hypothesis MCP (2HMCP) was proposed by analyzing the relationship between the error propagation effect and the hypothesis number [6]. It has been shown that a hypothesis number no larger than three is suitable for low bit rate video and MHMCP suppresses short term effect of error propagation more effectively than Intra refresh scheme. Ma et.al. extend the work in [6] to a more generalized one with reference frames being at arbitrary distances and analyze the error resilience characteristics of B pictures in H.264/AVC [7].

In this paper we improve the performance of MHMCP in packet loss scenarios by optimally estimating the overall distortion of the decoder at a pixel level. We use this estimate to make optimal mode decisions and optimize the prediction coefficients at the encoder so as to achieve the best possible end-to-end performance. This method extends the applicability of ROPE via a newly derived set of equations for the 2HMCP case, which is directly extendible to generalized B pictures,

MMSP'10, October 4-6, 2010, Saint-Malo, France.
978-1-4244-8112-5/10/$26.00 ©2010 IEEE.

and to motion compensation with more than 2 hypotheses.

II. ROPE ALGORITHM FOR 2HMCP

One of the key challenges of video communication over error prone channels is to mitigate the error propagation caused by the MCP scheme in a video codec. This can be achieved by having an appropriate amount of error resilience in the compressed bitstream. In this section we introduce the 2HMCP technique and derive equations to optimally estimate the total distortion of the decoder for a given packet loss condition and error concealment technique. The overall distortion is calculated at a pixel level via a simple recursion at the encoder. This ROPE estimate is then incorporated within the RD framework to optimally switch between Intra mode and 2HMCP mode for each MB and thus improving the overall performance and mitigating error propagation.

The prediction for a typical MHMCP scheme can be defined as

$$f_n^i = \sum_{k=1}^{m} \alpha_k \hat{f}_{n-k}^{g_k} \tag{1}$$

where f_n^i denotes the original value of pixel i in frame n, $\hat{f}_{n-k}^{g_k}$ is the motion compensated prediction from the k-th previous reconstructed frame, g_k denotes the position of the pixel due to displacement from MCP for hypothesis k and α_k is the corresponding prediction coefficient. The weights α_k satisfies the criteria $\sum_{k=1}^{m} \alpha_k = 1$ and $\alpha_k \geq 0$.

The 2HMCP scheme can thus be defined using (1) as

$$f_n^i = \alpha \hat{f}_{n-1}^a + (1-\alpha)\hat{f}_{n-2}^b \tag{2}$$

where the weights $\alpha_1 = \alpha$ and $\alpha_2 = (1-\alpha)$, the displacements $g_1 = a$ and $g_2 = b$.

A. Preliminaries

In our coding system the first frame is coded as an I frame and for the rest of the frames in the sequence the MBs are either coded in the 2HMCP mode (predictions are formed using linear combination of the best references from the two previous frames) according to equation (2) or coded as Intra MBs. We form a group of blocks (GOB) from all the MBs in a particular row (slice), and assume that each GOB is carried in a separate packet. In this setting, the loss rate of a pixel equals the packet loss rate p. We assume this value of p is available at the encoder.

The original frame is denoted by f_n and its encoder reconstruction as $\hat{f}_n$. When a packet is lost during transmission, the decoder uses an error concealment technique to estimate the missing video segment. Any error concealment method can be used and for simplicity we make use of temporal replacement method where the lost video segment is replaced by the co-located segment in the previous frame. Depending on the error concealment method, the ROPE equations has to be modified accordingly. We denote the decoded (and possibly error concealed) reconstruction of frame n as $\tilde{f}_n$. At the encoder we do not have access to this value and we therefore treat $\tilde{f}_n$ as a random variable.

At the decoder we make use of an error resilient technique, namely clean hypothesis formation [8]. This technique makes use of only the hypothesis from the corresponding reference frame that was received correctly. Consider an example where a MB in frame k is predicted from 2 hypotheses, one from frame $k-1$ and other from $k-2$. Suppose the hypothesis in $k-1$ was corrupt and concealed. At the decoder we form the prediction for the MB of frame k using entirely the hypothesis from frame $k-2$. This scheme has been shown to mitigate error propagation due to burst and random losses [8].

B. ROPE Algorithm

The overall expected mean-squared-error (MSE) distortion of a pixel is

$$\begin{aligned} E\{d_n^i\} &= E\{(f_n^i - \tilde{f}_n^i)^2\} \\ &= (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i\} + E\{(\tilde{f}_n^i)^2\} \end{aligned} \tag{3}$$

We observe that for calculating the overall distortion d_n^i we require the first and second moments of the random variable $\tilde{f}_n^i$. We develop recursive functions to calculate the two moments.

1) Pixel in an Intra coded MB:

$$E\{\tilde{f}_n^i\} = (1-p)\hat{f}_n^i + pE\{\tilde{f}_{n-1}^i\} \tag{4}$$

$$E\{(\tilde{f}_n^i)^2\} = (1-p)(\hat{f}_n^i)^2 + pE\{(\tilde{f}_{n-1}^i)^2\} \tag{5}$$

The recursive equations (4) and (5) are the same as derived in [1].

2) Pixel in a two hypothesis MB:

The prediction for pixel i in frame n using 2HMCP is as shown in (2). What is actually transmitted over the network is the compressed prediction error $\hat{e}_n^i$. This is given by

$$\hat{e}_n^i = \hat{f}_n^i - (\alpha \hat{f}_{n-1}^a + (1-\alpha)\hat{f}_{n-2}^b) \tag{6}$$

At the decoder if this packet is received correctly then it has access to both $\hat{e}_n^i$ and the displacement vectors. However, it has the decoder reconstruction of pixels a and b. Thus the decoder reconstruction of pixel i is given by

$$\tilde{f}_n^i = \hat{e}_n^i + (\alpha \tilde{f}_{n-1}^a + (1-\alpha)\tilde{f}_{n-2}^b) \tag{7}$$

At the encoder $\tilde{f}_{n-1}^a$ and $\tilde{f}_{n-2}^b$ are unknown and modeled as random variables. The probability of the packet correctly reaching the decoder is $1-p$ and if the packet is lost then it will be concealed using the pixel from the previous frame $\tilde{f}_{n-1}^i$. The first moment can be thus written as

$$E\{\tilde{f}_n^i\} = (1-p)[\hat{e}_n^i + E\{h_n^i\}] + pE\{\tilde{f}_{n-1}^i\} \tag{8}$$

where $E\{h_n^i\}$ is the expected value for the prediction from the two hypotheses. The probability that the packets containing both hypotheses were correctly received is $(1-p)^2$. The probability that one of the hypotheses was corrupted and the other was received correctly is $(1-p)p$. The probability that both the hypotheses were corrupt is p^2. When one of the hypotheses is corrupt, the decoder makes use of the correct hypothesis completely in forming the prediction as discussed

above and when both hypotheses are corrupt then it performs concealment by using the pixel from the previous frame $\tilde{f}^i_{n-1}$. Thus $E\{h^i_n\}$ can be defined as

$$\begin{aligned} E\{h^i_n\} &= (1-p)^2[\alpha E\{\tilde{f}^a_{n-1}\} + (1-\alpha)E\{\tilde{f}^b_{n-2}\}] \\ &\quad +(1-p)pE\{\tilde{f}^a_{n-1}\} + p(1-p)E\{\tilde{f}^b_{n-2}\} \\ &\quad +p^2E\{\tilde{f}^i_{n-1}\} \end{aligned} \tag{9}$$

The second moment of $\tilde{f}^i_n$ is given by

$$\begin{aligned} E\{(\tilde{f}^i_n)^2\} &= (1-p)E\{[\hat{e}^i_n + E\{h^i_n\}]^2\} + pE\{(\tilde{f}^i_{n-1})^2\} \\ &= (1-p)[(\hat{e}^i_n)^2 + 2\hat{e}^i_n E\{h^i_n\} + E\{(h^i_n)^2\}] \\ &\quad +pE\{(\tilde{f}^i_{n-1})^2\} \end{aligned} \tag{10}$$

The second moment for the prediction from the two hypothesis can be written as

$$\begin{aligned} E\{(h^i_n)^2\} &= (1-p)^2E\{[\alpha\tilde{f}^a_{n-1} + (1-\alpha)\tilde{f}^b_{n-2}]^2\} \\ &\quad +(1-p)pE\{(\tilde{f}^a_{n-1})^2\} + p(1-p) \\ &\quad E\{(\tilde{f}^b_{n-2})^2\} + p^2E\{(\tilde{f}^i_{n-1})^2\} \end{aligned} \tag{11}$$

The first term in (11) can be expanded as

$$\begin{aligned} E\{[\alpha\tilde{f}^a_{n-1} + (1-\alpha)\tilde{f}^b_{n-2}]^2\} = \alpha^2E\{(\tilde{f}^a_{n-1})^2\} \\ +2\alpha(1-\alpha)E\{\tilde{f}^a_{n-1}\tilde{f}^b_{n-2}\} + (1-\alpha)^2E\{(\tilde{f}^b_{n-2})^2\} \end{aligned} \tag{12}$$

where $E\{\tilde{f}^a_{n-1}\tilde{f}^b_{n-2}\}$ is the cross correlation term.

1) Cross Correlation Model: We use a simple approximation and assume the two pixels $\tilde{f}^a_{n-1}$ and $\tilde{f}^b_{n-2}$ are correlated when they have the same motion vectors (i.e. $a = b$) and we assume the two pixels are uncorrelated when they are pointing to different spatial locations in the frame (i.e. $a \neq b$). For the case where we assume the pixels are correlated we use a linear signal model proposed in [9] to account for the correlation. More sophisticated approximations as those in [9] are possible and developing such models that capture the current setting may improve results further but are left for future work.

We reemphasize that the recursive equations derived in (8) and (10) for first and second moments are substituted in equation (3) in order to calculate the expected distortion at the decoder. The encoder exploits this result directly to select both modes and prediction coefficients to optimize the rate distortion tradeoff.

III. Simulation Results

We implemented the 2HMCP scheme on top of JM 15.1 reference software [10]. All sequences were encoded at 15fps, QCIF resolution. We adopted the rate control from the JM codec and set one common quantization scale for all the MBs of a frame. The first frame was coded as an Intra frame and the rest of the frames in the sequence were coded in 2HMCP mode; i.e. predictions are formed using a linear combination of the best references from the two previous frames using equation (2). The MBs in a 2HMCP frame were coded either in 2HMCP mode or as an Intra MB. The second frame in the sequence is coded as a P frame, as it has only one reference frame. The mode decision for the MBs in 2HMCP frame depends on the source coding distortion in the standard 2HMCP scheme ("std_2HMCP") and end-to-end distortion in our proposed scheme ("ROPE_2HMCP"). The prediction coefficients have been optimized using the end-to-end distortion for our proposed scheme. At the decoder for each packet loss rate (PLR), 200 randomly generated packet loss patterns were applied, and the average luminance PSNR was computed to measure the system performance. For comparison we also provide results for conventional IPPP... scheme ("std_IPPP") where the first frame is coded as an Intra frame, the rest as P-frames and ROPE optimal MB coding mode selection for the IPPP... scheme ("ROPE_IPPP") as proposed in [1].

A. Performance vs. Bit-rate

In the first experiment we compared the three methods (std_IPPP, std_2HMCP and ROPE_2HMCP) across a range of effective bit-rates with the packet loss rate fixed at 10%. Figure 1 shows the results for Foreman and Stefan test sequences. We observe that std_2HMCP scheme performs better than std_IPPP scheme proving the better error resilience property of 2HMCP. ROPE_2HMCP scheme outperforms std_2HMCP scheme by around 3-5 dB due to the effectiveness of the optimal mode decision.

B. Performance vs. PLR

In the second experiment we compared the three techniques across a PLR range of 1% to 20% at a fixed bit-rate of 144 kbps. Figure 2 shows the results for Foreman and Stefan test sequences. We observe that ROPE_2HMCP scheme performs better than all the competing techniques across different PLRs. This demonstrates the effectiveness of using accurate end-to-end distortion estimates in mode decision.

In Figure 3 we compare the ROPE techniques (ROPE_IPPP and ROPE_2HMCP) for Foreman sequence at 15 fps. We observe that conventional ROPE (ROPE_IPPP) already eliminates a substantial amount of damage due to loss and error propagation. However we observe that ROPE_2HMCP scheme (using the simple cross correlation model explained above) has moderate gains ranging from 0.4 to 0.8 dB over conventional ROPE. This indicates the promise of 2HMCP scheme with efficient end-to-end distortion estimation.

IV. Conclusion and Future work

Besides good compression efficiency offered by MHMCP, it has been proven to have a good error resilience property. In this paper we proposed an effective mode switching algorithm for the 2HMCP scheme adopting ROPE to compute the end-to-end distortion in packet loss scenarios. This improvement in the overall performance provides a good reproduction video quality during transmission of video over unreliable networks.

The proposed scheme can be further enhanced by developing more sophisticated models for the cross correlation term.

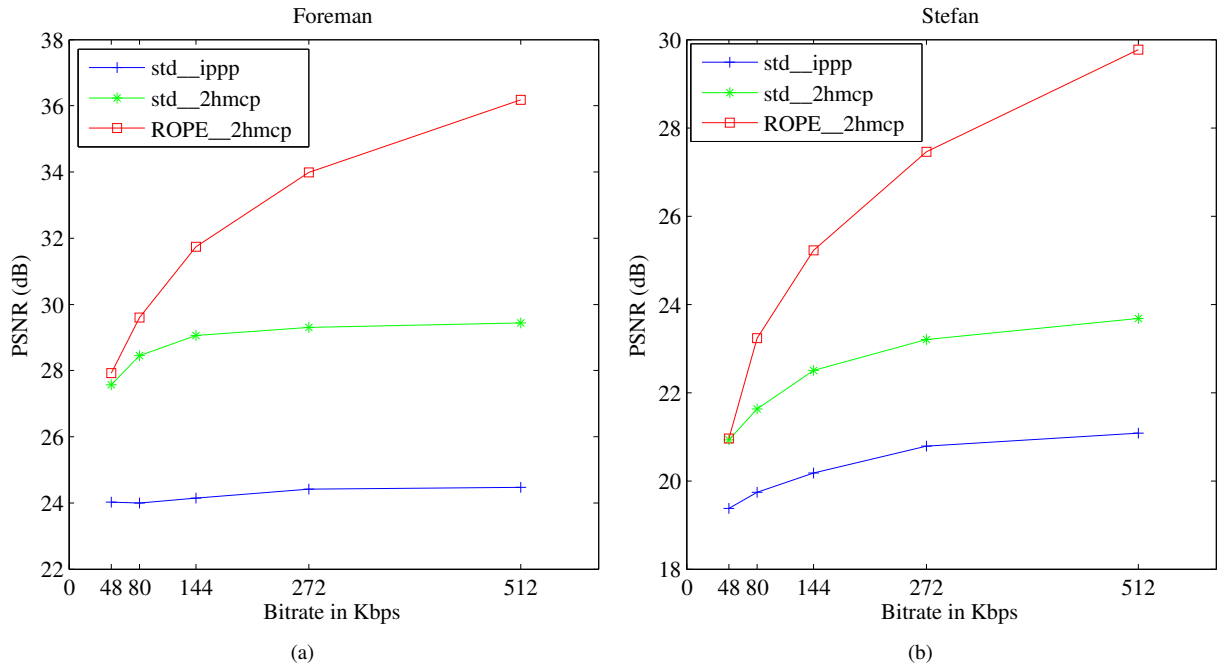

Fig. 1. PSNR Vs Bit-rate curves for (a) Foreman and (b) Stefan sequences at 10% PLR for all competing techniques.

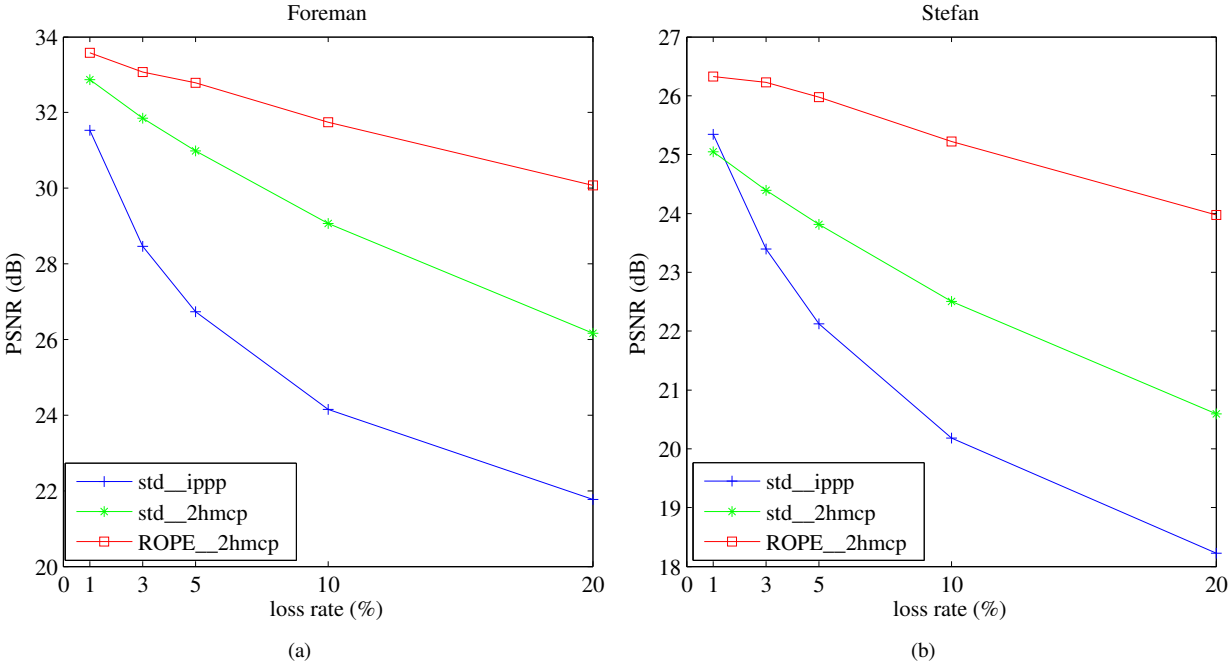

Fig. 2. PSNR Vs packet loss curves for (a) Foreman and (b) Stefan sequences at 15 fps and 144 kbps for all competing techniques.

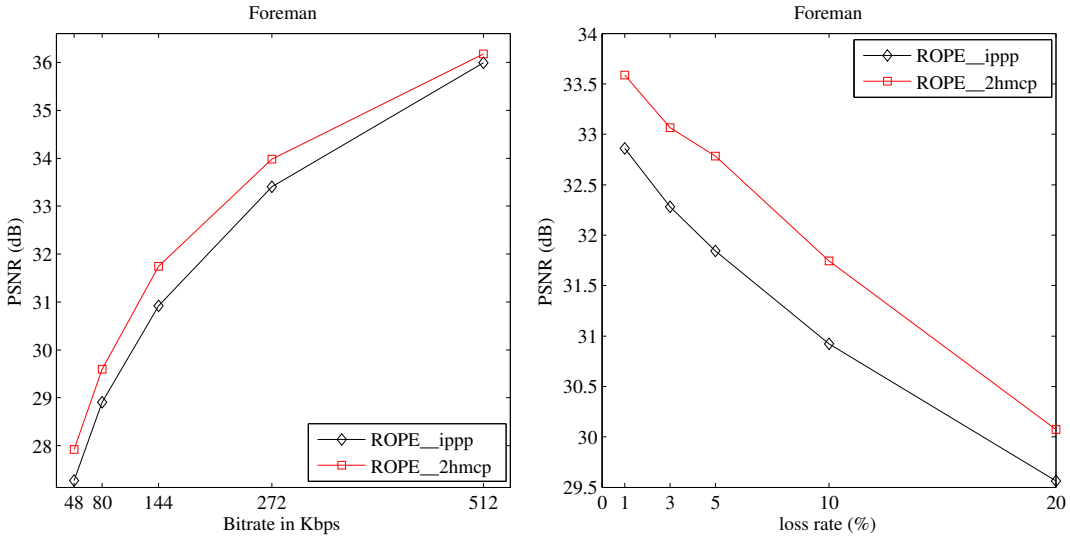

Fig. 3. (a) PSNR VS Bit-rate at 10% PLR (b) PSNR VS PLR at 144 kbps curves for Foreman sequence at 15 fps for ROPE techniques.

REFERENCES

[1] R. Zhang, S. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966–976, June 2000.

[2] A. Leontaris and P. Cosman, "Video compression for lossy packet networks with mode switching and a dual frame buffer," *IEEE Trans. Image processing*, vol. 13, no. 7, pp. 885–897, July 2004.

[3] M. Budagavi and J. D. Gibson, "Multiframe block motion compensated video coding for wireless channels," *in Proc. 30th Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 953 – 957, Nov 1996.

[4] B. Girod, "Efficiency analysis of multihypothesis motion-compensated prediction for video coding," *IEEE Trans. Image processing*, vol. 9, no. 2, pp. 173–183, Feb. 2000.

[5] S. Lin and Y. Wang, "Error resilience property of multihypothesis motion-compensated prediction," *in Proc. IEEE Intl Conf. Image Processing*, vol. 3, pp. 545–548, June 2002.

[6] W.-Y. Kung, C.-S. Kim, and C.-C. Kuo, "Analysis of multihypothesis motion-compensated prediction (MHMCP) for robust visual communication," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 146–153, Jan. 2006.

[7] M. Ma, O. Au, S.-H. Chan, and L. Guo, "Error resilient video coding using B pictures in H.264," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 10, pp. 1448–1461, Oct. 2009.

[8] Y.-C. Tsai and C.-W. Lin, "H.264 error resilience coding based on multihypothesis motion-compensated prediction," *in Proc. IEEE Intl Conf. Multimedia and Expo*, pp. 952 – 955, July 2005.

[9] 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. Circuits Syst. Video Technol.*, vol. 17, no. 7, pp. 845–856, July 2007.

[10] "Joint Video Team (JVT) Reference Software." [Online]. Available: http://iphome.hhi.de/suehring/tml/