ERROR RESILIENCE PROPERTY OF MULTIHYPOTHESIS MOTION-COMPENSATED PREDICTION *

Shunan Lin and Yao Wang
Department of Electrical and Computer Enginnering
Polytechnic University, Brooklyn, NY 11201, USA
{slin,yao}@vision.poly.edu

Abstract - Multihypothesis motion-compensated prediction (MHMCP) approach has shown significant gain in terms of coding efficiency both in theory and practice. However, the fact that it also enhances the error resilience of a codec has not been fully realized. This paper analyzes the error resilience gain of MHMCP. Models for the decoder distortion induced by transmission errors and the rate-distortion function of the encoder in a codec using MHMCP are presented and their accuracy is validated by simulation studies. Using these models, we show how to design the multihypothesis predictor to jointly consider the coding gain and the error resilience gain for given channel error characteristics.

INTRODUCTION

Today's state-of-the-art video techniques incorporate motion-compensated prediction (MCP) to exploit the temporal correlations of video signals. Some of these techniques, for example, B-frames, employ multihypothesis motion compensation approach in which a linear superposition of more than one MCP signals is used to predict the current macroblock. Multihypothisis MCP (MHMCP) has shown significant coding gains in combination with multiple reference pictures both in theory and in practice [1, 2, 3].

Although the original objective of multihypothesis pictures is to improve the coding efficiency, it can also improve the robustness of coded bit streams. The underlying reason of the error resilience gain in a MHMCP system is that the linear combination of prediction signals acts as a loop filter, which is a well-known technique to increase the robustness of DPCM system by suppressing the error propagation [4, 5, 6]. The advantages of using MHMCP are analogous with half-pixel motion estimation: both can improve coding efficiency

*THIS WORK WAS SUPPORTED IN PART BY NSF GRANT ANI 0081357 AND BY THE NEW YORK STATE CENTER FOR ADVANCED TECHNOLOGY IN TELECOMMUNICATION (CATT).

and both have the effect of mitigating error propagation. In practice, error resilience gain by MHMCP was recently reported in several papers. In [7], a double-vector motion compensation technique, which essentially is the same as two-hypothesis prediction except that the hypothesis is limited to the previous two frames, was proposed to improve the robustness of the coded streams. In [8], we proposed a coding scheme which also uses two-hypothesis prediction to enhance error resilience. Although our idea is originally developed for multiple description coding, it is found to be also applicable in traditional single description coding.

A critical issue in designing a MHMCP codec is what predictor to use. The optimal linear predictor was derived in terms of coding efficiency in [1]. In [7], the influence of the predictor on the error propagation in the proposed algorithm was discussed. The discussion, however, did not consider the coding efficiency. Furthermore, it was only developed for that specific algorithm and gave only an upper bound for the propagated error, which cannot capture the dynamic behavior of the errors.

We argue that the predictor in a MHMCP codec for video transmission over error prone channels should be designed to achieve a good trade-off between error resilience and coding efficiency. In this paper, we develop a model for the decoder distortion due to transmission errors, and a model for the encoder R-D performance, both for a codec using two-hypothesis prediction. Using these models, the optimal predictor can be derived numerically for a given channel loss rate.

The paper is organized as follows: In section 2, we describe the codec using MHMCP and derive a model for the distortion induced by transmission errors. A simple model for the encoder's R-D performance is given in section 3. In section 4 the developed models are tested for video transmission over error prone channels. Conclusions are given in section 5.

MULTIHYPOTHESIS PICTURES AND DECODER DISTORTION MODEL FOR TWO-HYPOTHESIS PICTURES

A compressed video signal is vulnerable to transmission errors. One of the key challenges in combating video transmission errors is how to mitigate error propagations caused by the MCP scheme in a video codec. In this section, we will give the concept of MHMCP and discuss how the MHMCP can help the decoder reduce the error propagations.

In this paper, the MHMCP technique we consider is only for multihypothesis pictures, as defined in [9], which are pictures whose macroblocks are compensated by a linear combination of motion-compensated macroblocks. They are like B-pictures except that their reference frames are temporally previous pictures. Because of such a difference, there is no extra decoding delay and multihypothesis pictures can also be used as reference frames.

A general multihypothesis picture $\psi(n)$ can be predicted by

$$\hat{\psi}(n) = \sum_{k=1}^{m} h_k \tilde{\psi}_r(n-k). \tag{1}$$

with $\tilde{\psi}_r(n-k)$ being a motion-compensated prediction from the k-th previous reconstructed picture and h_k being the corresponding predictor coefficient. The set $\{h_k\}$ satisfy $\sum_{k=1}^{m} h_k = 1$ and $h_k \geq 0,\ k = 1, 2, 3 \ldots$. To make the notations simple, the spatial locations of the signals are ignored. Notice if h = [1,0,0,..], the predictor becomes conventional "single hypothesis".

Let us first discuss the error propagation in a "single hypothesis" codec. Assume the information of frame $\psi(n)$ is lost during transmission. The error between the decoded frame $\psi_d(n)$ and the encoded frame $\psi_e(n)$, which depends on the error concealment method, is given by

$$\epsilon_{de}(n) = \psi_d(n) - \psi_e(n). \tag{2}$$

If the information for all the remaining frames are received correctly, and if we don't consider the other mismatch mitigation methods such as intra update or explicit and implicit spatial loop filters, it is easily shown that the mismatch error for frames $n+k$ is the same as that for frame n, i.e.

$$\epsilon_{de}(n+k) = \epsilon_{de}(n). \qquad k = 1, 2, \ldots. \tag{3}$$

This means that there is no error suppression effect in a conventional MCP scheme. In a multihypothesis codec, the mismatch error for $\psi(n)$ is the same as in (2) if the previous frames are transmission error-free and the error concealment method for frame $\psi(n)$ is identical to that in the single hypothesis codec. However, the errors for the following frames are different:

$$\epsilon_{de}(n+1) = h_1 \epsilon_{de}(n), \tag{4}$$

$$\epsilon_{de}(n+2) = h_1 \epsilon_{de}(n+1) + h_2 \epsilon_{de}(n) = (h_1^2 + h_2)\epsilon_{de}(n), \tag{5}$$

$$\ldots\ldots$$

If $h_k > 0,\ k = 1, 2, 3 \ldots$, it is easily shown that $|\epsilon_{de}(n+k)| < |\epsilon_{de}(n)|.\ k = 1, 2, 3 \ldots$. So the multihypothesis picture can recover from the errors more quickly than conventional single hypothesis codec. Furthermore, from above equations, it is obvious that the error resilience gain depends on the predictor $\{h_k\}$ used. Notice the selection of the predictor also influence the coding efficiency. Hence the predictor should be designed to achieve a good trade-off between error robustness and coding efficiency.

In general, given a video signal, the coding efficiency theory of a MHMCP codec depends on the number of the hypothesis, the predictor parameters and the motion estimation algorithm. All of the above factors are intensively investigated in [1, 10]. The focus of this paper, however, is not the coding efficiency but the error resilience property of MHMCP. So, in the rest of this paper, we assume the codec uses only two hypotheses, which means previous two frames are used as references, and the whole sequence except I-frames are encoded as multihypothesis pictures. The influence of the predictor on the coding efficiency and the error resilience will be studied in this kind of codec.

Assume the information of frame $\psi(n)$ is lost during transmission and the other information is received correctly. Then we have

$$\begin{aligned}
\epsilon_{de}(n) &= \psi_d(n) - \psi_e(n) = \epsilon_0, && (6)\\
\epsilon_{de}(n+1) &= h_1\epsilon_0, && (7)\\
\epsilon_{de}(n+2) &= h_1\epsilon_{de}(n+1) + h_2\epsilon_{de}(n) = (h_1^2 + h_2)\epsilon_0, && (8)\\
&\cdots\cdots \\
\epsilon_{de}(n+k) &= h_1\epsilon_{de}(n+k-1) + h_2\epsilon_{de}(n+k-2). && (9)
\end{aligned}$$

From above equations, we can obtain

$$\epsilon_{de}(n+k) = \frac{1-(h_1-1)^{k+1}}{2-h_1}\epsilon_0. \tag{10}$$

The function in (10) is an oscillation function with a decreasing envelop. The smaller h_1 is, the smaller the error converges to, but at the expense of more severe oscillation. Note that (7) and (8) assume the lost frame is still used in decoding the frames followed. The results and conclusions in this section and the following sections are based on this decoding assumption.

Above error propagation analysis does not take into account the effects of the spatial filtering due to the use of half-pixel motion-compensated prediction. In [6], the error attenuation effects of the spatial filtering is modeled by

$$\epsilon_{de}^2(n+k) = \frac{\epsilon_{de}^2(n)}{1+\gamma_d k} = \frac{\epsilon_0^2}{1+\gamma_d k}, \tag{11}$$

where, to make the notation simple, we use $\epsilon_{de}^2(n+k)$ to denote the mean square error (MSE) for frame $n+k$ and ϵ_0^2 to denote the variance of the samples

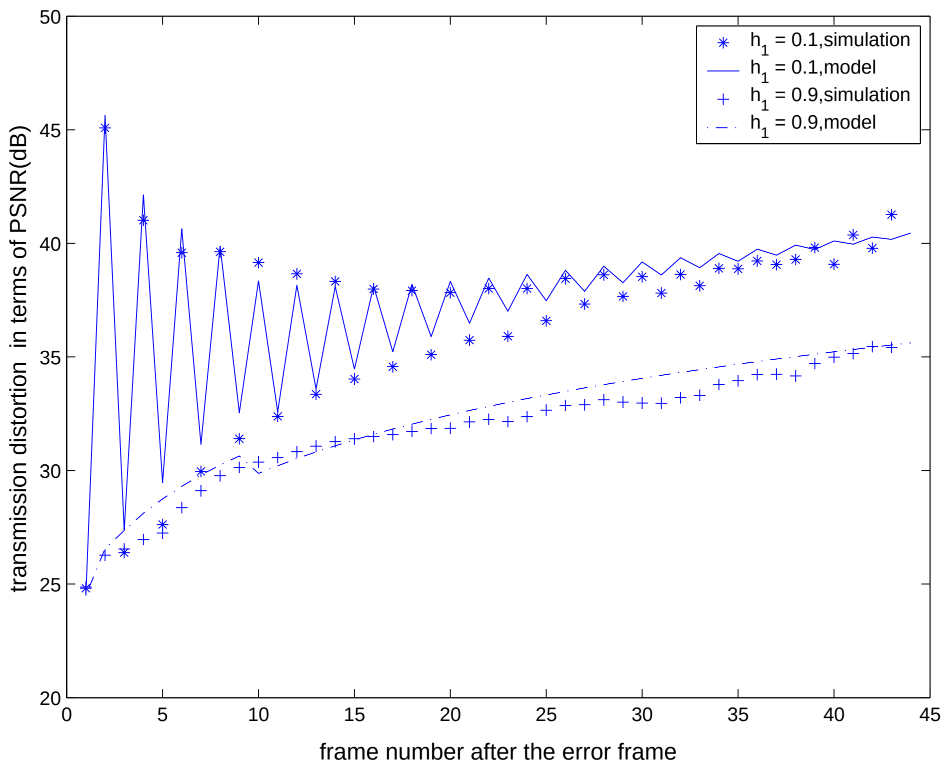

Figure 1: Decoder model verification: γ_d is 0.3.

in the error frame ϵ_0. Combining (10) and (11) generates

$$\epsilon_{de}^2(n+k) = \left(\frac{1-(h_1-1)^{k+1}}{2-h_1}\right)^2 \frac{\theta}{1+\gamma_d k}\,\epsilon_0^2. \tag{12}$$

The new parameter θ is a constant larger than 1, which is due to the fact that the strength of the decay from (11) is reduced a little by the multihypothesis. In our experiments, we found it is about 1.3.

Figure 1 displays the error propagation decay effect in one of our simulations. The QCIF video sequence 'Foreman" (frames 1-200) is coded at 15fps and 128kbps. The codec is h.263+ codec from [11]. TMN8.0 rate control method is used but the frame layer rate control is disabled. Two motion vectors for two hypotheses are estimated separately since our focus is not coding efficiency. No other options are used except annex E (arithmetic coding). The copy-previous-frame method is used for error concealment. In figure 1 one frame in the sequence is randomly chosen to be lost and the transmission distortion after that frame is plotted. The distortion estimated by (12) is also shown. From this figure, we can see that the model describes the decoder behavior very accurately. The oscillation phenomenon is also clearly observed when h_1 is small, as can be expected from (9). Overall, a small h_1 is superior to a large h_1 in terms of error robustness.

To compute time average distortion per frame $\bar{\epsilon^2}$, let us assume an I-frame is encoded every N frames. Then the error propagation extends at most over N frames. Assume the frame loss rate is P. For frame n, the probability that its k-th previous frame is lost is P and its effect on this frame is, according to (12),

$$E(k) = \epsilon_0^2((1-(h_1-1)^{k+1})^2\theta)/((2-h_1)^2(1+\gamma_d k)).$$

The average distortion for this frame is thus

$$\bar{\epsilon^2_{de}}(n) = P\sum_{k=0}^{n} E(k), n = 0, 1, \ldots, N-1. \quad (13)$$

Hence the average distortion is:

$$\bar{\epsilon^2_{de}} = \frac{1}{N}\sum_{n=0}^{N-1}\bar{\epsilon^2_{de}}(n) = \frac{P}{N}\sum_{k=0}^{N-1}(N-k)E(k). \quad (14)$$

The verification of this average distortion model will be shown in section 4.

ENCODER DISTORTION MODEL FOR TWO-HYPOTHESIS PICTURES

In the last section, we noted that the selection of the predictor should jointly consider the coding performance and transmission robustness. Therefore, in addition to the model for transmission distortions, we need to establish a model for the predictor's influence on the R-D performance of the encoder. Although a general model was given in [1], here we can use a simpler model since we only have two hypotheses.

Assume the predictor is h_1 and $h_2 = 1 - h_1$, the mean square prediction error is

$$\begin{aligned} \sigma_p^2 &= E\{(\psi(n) - h_1\tilde{\psi}_e(n-1) - h_2\tilde{\psi}_e(n-2))^2\} \\ &= 2\sigma^2(1-\beta-(1+\alpha-\beta-\gamma)h_1 + (1-\gamma)h_1^2). \end{aligned} \quad (15)$$

in which $\sigma^2 = E\{\psi^2(n)\} = E\{\psi_e^2(n)\}$, $\alpha = E\{\psi(n)\tilde{\psi}_e(n-1)\}/\sigma^2$, $\beta = E\{\psi(n)\tilde{\psi}_e(n-2)\}/\sigma^2$ and $\gamma = E\{\tilde{\psi}_e(n-1)\tilde{\psi}_e(n-2)\}/\sigma^2$. To find a simple

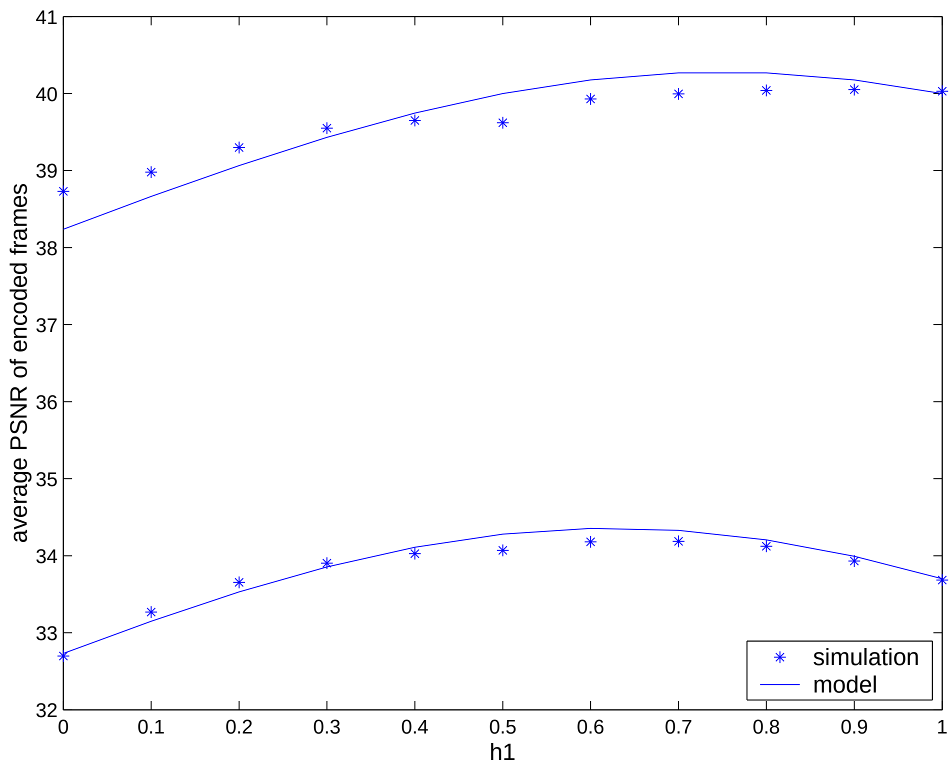

Figure 2: Encoder model verification: The parameters α, β and γ for sequences "mother and Daughter" and "Foreman" are $\{0.990, 0.985, 0.990\}$ and $\{0.980, 0.975, 0.980\}$ respectively.

rate distortion model for the video encoder, we assume the encoding distortion is proportional to the prediction errors, i.e. $\sigma_e^2 \propto \sigma_p^2$. Then, the encoding distortion can be expressed as

$$\begin{aligned} \epsilon_e^2 &= \varepsilon^2 \sigma_p^2 \\ &= \varepsilon^2 \sigma^2 (1 - \beta - (1 + \alpha - \beta - \gamma) h_1 + (1 - \gamma) h_1^2), \end{aligned} \tag{16}$$

where ε^2 represents the factors that do not change with the predictors (it depends on the bit rate and the probability distribution of the prediction error). From (16), we can see that the influence of the predictor on the coding efficiency depends on the correlation coefficients α, β, γ, which in turn depend on the similarity of the successive frames and the motion estimation accuracy.

Computer simulations are conducted to verify this model. The experiment condition is the same as used in the last section. To see the effect that different signals prefer different predictors, a lower motion sequence "Mother and Daughter" is also coded at the same frame rate and bit rate as "Foreman". The average PSNRs for frames 1-200 for different predictors are given in figure 2. When calculating the distortion according to the model, we set ε^2 according to the simulation results of "single hypothesis", i.e., when $h = \{1, 0\}$. The correlation coefficients α, β and γ are determined from the actual coded sequence. From figure 2, we can see the model matches the test results quite closely. Also, we can see "Mother and Daughter" has a larger optimal h_1 than "Foreman".

SIMULATION OF CODED VIDEO OVER ERROR PRONE CHANNELS

Having both the encoder distortion function and the transmission distortion function, we can express the total distortion in the decoder as

$$D = \epsilon_e^2 + \epsilon_{de}^{\bar{2}}. \tag{17}$$

For a given sequence and frame loss rate P, the optimal predictor can be found numerically by finding h_1 that minimizes the total distortion D. In this section, we present experiment results to verify the accuracy of the above models.

Sequence "Foreman" is encoded as explained in section 2. The encoded bit stream is transmitted through a packet lossy channel. We assume each packet contains the data from an entire frame and the packet loss rate (equals to the frame loss rate) is set at $P = 1\%, 3\%, 5\%$ respectively. The encoded stream is transmitted 30 times and the frames are randomly dropped according to the specified loss rate P. The average PSNR using different predictors under different packet loss rates are plotted in Figure 3. From this figure, it can be seen i) there is 1-3dB performance differences for different predictors. The worse the channel is, the higher difference it makes; ii) the optimal predictor for coding efficiency, which is $h_1 = 0.7$ according to figure 2, has 0.2-1.5dB performance loss compared with the jointly optimal predictors; iii) when the channel is bad, the robustness is the dominant factor determining the total distortion, so the optimal h_1 goes to a small number; when the channel is good, the coding efficiency becomes more important and the optimal h_1 is larger. Also, we can see the model developed is accurate and the optimal predictor determined from the model curve is quite close to that determined from the experimental data. Note that the experiments and the above observations are based on the simple copy-previous-frame error concealment technique. If more sophisticated error concealment technique is used, all formulas are still valid although the value of optimal predictor might be different.

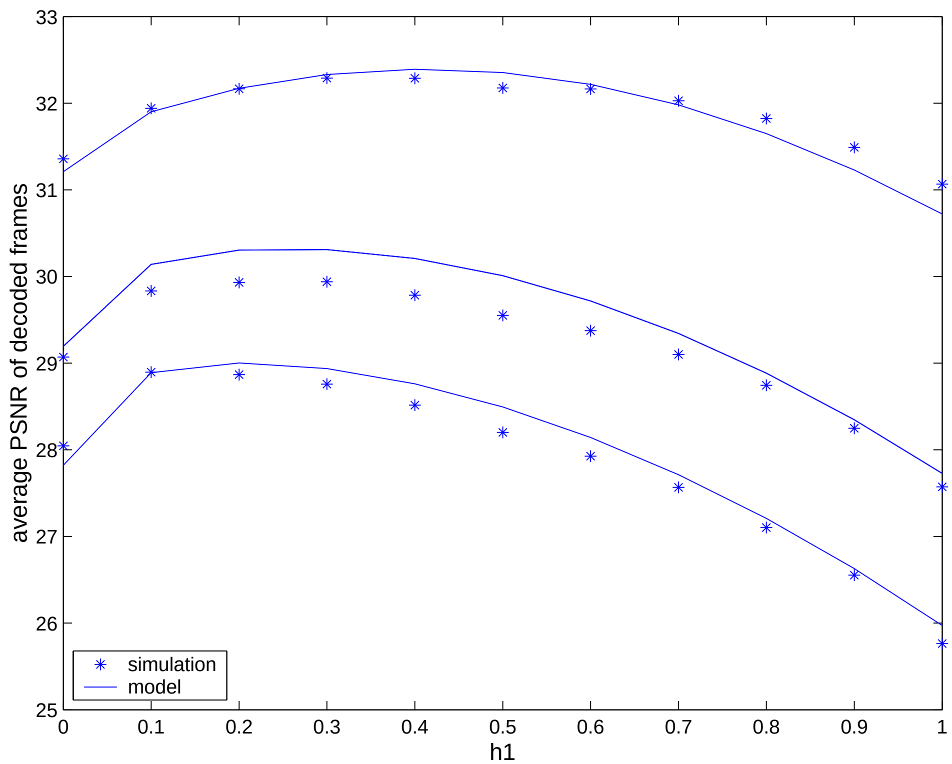

Figure 3: Total distortion for “Foreman” under different packet loss rates.

CONCLUSION

Multihypothesis motion-compensated prediction can improve robustness as well as efficiency of a video codec for video transmission. The predictor selection in a MHMCP codec should jointly consider coding efficiency and error resilience. Two models were developed which can measure accurately how the predictors influence transmission distortion and R-D performance of a two-hypothesis codec. The optimal predictor can be obtained numerically based on these models, given the bit rate and the channel packet loss rate.

References

[1] B. Girod, “Efficiency Analysis of Multihypothesis Motion-Compensated Prediction for Video Coding,” *IEEE Transactions on Image Processing*, vol. 9, no. 2, pp. 173-183, Feb. 2000.

[2] M. Flierl, T. Wiegand, and B. Girod, “Rate-Constrained Multi-Hypothesis Motion-Compensated Prediction for Video Coding,” in *Proc.ICIP*, (Vancouver, Canada), Sept. 2000, pp. 150-153.

[3] M. Flierl and B. Girod, “Further Investigation of Multihypothesis Motion Pictures,” ITU-T SG16/Q6 VCEG-M40, Austin, Texas, USA, April 2001

[4] D.J.Connor, “Techniques of Reducing the Visibility of Transmission Errors in Digitally Encoded Video Signals,” *IEEE Transactions on Communications*, vol. COM-21, no. 6, pp. 695 - 706, Jun. 1973.

[5] P.Haskell and D. Messerschmitt, “Resynchronization of motion compensated video affected by ATM cell loss,” in *Proc. ICASSP*, (Sanfrancisco USA), Mar. 1992, pp. 545-548.

[6] B. Girod and N. Färber, ”Wireless Video,” in *Compressed Video over Networks*, A. Reibman, M.-T. Sun (eds.), Marcel Dekker, 2000.

[7] C.S. Kim, R.C. Kim and S.U. Lee, “Robust Transmission of VIdeo Sequence Using Double-Vector Motion Compensation,” *IEEE Transactions on Circuits ans Systems for Video Technology*, vol. 11, no. 9, pp. 1011 - 1021, Sep. 2001.

[8] Y. Wang and S. Lin, “ Error Resilient Video Coding using Multiple Description Motion Compensation,” in *Proc. MMSP*, (Cannes, France), Oct. 2001, pp. 441-446.

[9] M. Flierl and B. Girod, “ Multihypothesis Pictures for H.26L,” in *Proc. ICIP*, (Thessaloniki, Greece) , Oct. 2001.

[10] M. Flierl, T. Wiegand and B. Girod, “A Video Codec Incorporating Block-Based Multi-Hypothesis Motion-Compensated Prediction,” in *Proc. SPIE Conference on Visual Communications and Image Processing*, (Perth, Australia), Jun. 2000.

[11] University of British Columbia, “H.263+ codec”, ftp://dspftp.ece.ubc.ca.