RATE-CONSTRAINED MULTI-HYPOTHESIS MOTION-COMPENSATED PREDICTION FOR VIDEO CODING

Markus Flierl, Thomas Wiegand

Telecommunications Laboratory
University of Erlangen-Nuremberg
Erlangen, Germany
{flierl, wiegand}@LNT.de

Bernd Girod

Information Systems Laboratory
Stanford University
Stanford, CA
girod@ee.stanford.edu

ABSTRACT

Multi-hypothesis prediction extends motion compensation with one prediction signal to the linear superposition of several motion-compensated prediction signals with the result of increased coding efficiency. The multiple hypotheses in this paper are blocks in past decoded frames. These blocks are referenced by individual motion vectors and picture reference parameters incorporating long-term memory motion-compensated prediction. In this work, we at most employ two hypotheses similar to B-frames. However, they are obtained from the past. Due to the increased rate for the motion vectors, rate-constrained coder control is utilized. For this scheme, we demonstrate the coding efficiency of multi-hypothesis prediction in combination with variable block size and long-term memory and present bit-rate savings up to 32% when compared to standard variable block size prediction without long-term memory motion compensation.

1. INTRODUCTION

Today's hybrid video coding schemes use successfully block-based motion-compensated prediction (MCP). It is well known that the achievable MCP performance can be increased by reducing the size of the motion-compensated blocks [1]. Additional improvements can be obtained by long-term memory MCP. This concept increases the number of reference frames available for MCP [2].

Many of these video coding schemes employ more than one MCP signal simultaneously. The term "multi-hypothesis motion compensation" has been coined for this approach. A linear combination of multiple prediction hypotheses is formed to arrive at the actual prediction signal. The efficiency of multi-hypothesis MCP for video coding is analyzed in [3].

B-Frames, as they are standardized in H.263 [4] or MPEG, are an example of multi-hypothesis motion compensation where two motion-compensated signals are superimposed to reduce the bit-rate of a video codec. But the B-Frame concept has to deal with a significant drawback: prediction uses the reference pictures before and after the B-picture. The associated delay may be unacceptable for interactive applications. To overcome this disadvantage the authors proposed rate-constrained prediction algorithms in [5, 6] which benefit from the idea of superimposing prediction signals, but select them from the past frames only.

The authors presented in [7] a video codec that incorporates multi-hypothesis motion-compensated prediction as proposed in [5] and showed that two jointly optimized hypotheses are efficient for practical video compression algorithms.

In this paper, we employ just two jointly optimized hypotheses and examine the influence of long-term memory and variable block sizes on block-based multi-hypothesis prediction and demonstrate the efficiency of the combination of the three concepts.

2. MULTI-HYPOTHESIS VIDEO CODEC

The multi-hypothesis video codec incorporates long-term memory motion compensation, which improves the efficiency of motion compensation (MC) by adding a frame reference parameter to each motion vector. This permits the use of several decoded frames instead of only the previously decoded picture for block-based MC. Fig. 1-a depicts the concept of long-term memory MC as published in [2].

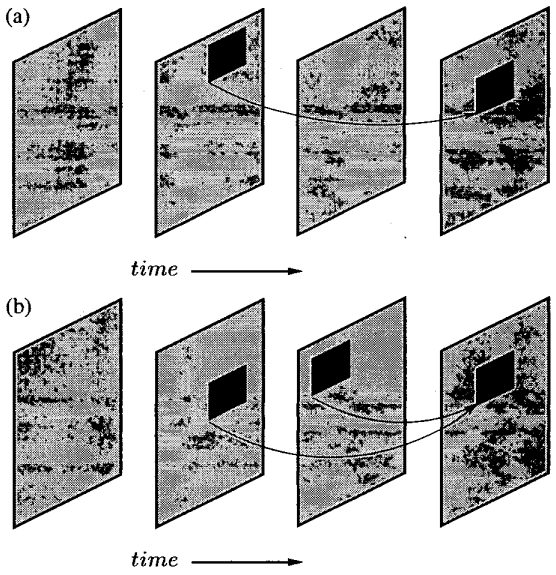

Fig. 1. (a) Long-term memory motion compensation. A block of a previous decoded frame is a prediction signal for the current frame. (b) Multi-hypothesis long-term memory motion compensation. Two blocks of previous decoded frames are linearly combined to form a prediction signal for the current frame.

0-7803-6297-7/00/$10.00 © 2000 IEEE

The long-term memory motion compensator is the basic unit providing the signals to be superimposed by our multi-hypothesis predictor. Let us consider two prediction signals – two hypotheses – $\mathbf{c}_1$ and $\mathbf{c}_2$ generated by long-term memory MC according to Fig. 1-b. The multi-hypothesis prediction signal $\hat{\mathbf{s}}$ is the superposition of these two prediction signals. Before adding the signals we weight each hypothesis by a constant coefficient. Results on the design of block-based multi-hypothesis motion-compensating predictors suggest that averaging the prediction signals with

$$\hat{\mathbf{s}} = \frac{1}{2}\mathbf{c}_1 + \frac{1}{2}\mathbf{c}_2 \tag{1}$$

is efficient [5].

A model analysis in [7] showed that prediction performance of two jointly optimized hypotheses is close the theoretical limit of multi-hypothesis motion-compensated prediction with an infinite number of averaged hypotheses. Experimental results in [7] also confirmed that two jointly optimized hypotheses are efficient for practical video compression algorithms as they provide a good trade-off between coder complexity and compression efficiency.

By employing at most two hypotheses, we have permitted the coding of a block in the current frame by two motion vectors and two frame reference parameters. The transmission of the second motion vector and frame reference parameter for each block potentially increases the bit-rate, which has to be justified by improved MCP. This trade-off limits the efficiency of the proposed approach. Improved efficiency can be obtained by adaptively switching between multi-hypothesis prediction and long-term memory prediction. Long-term memory prediction is long-term memory multi-hypothesis prediction with just one hypothesis.

State-of-the-art codecs employ variable block size MCP, for example, the advanced prediction mode described in the ITU-T Recommendation H.263 [4]. In a well-designed video coding scheme the most efficient concepts should be combined. Therefore, we apply the concept of adaptive multi-hypothesis prediction to macroblocks of size 16 × 16 as well as to blocks of size 8 × 8.

The syntax of the H.263 code is extended so that adaptive multi-hypothesis motion compensation is possible. On the macroblock level, we add the new INTER2H code. This code is similar to the INTER code of H.263. The new mode additionally codes for the second hypothesis an extra motion vector and frame reference parameter. For advanced prediction, the INTER4V code is extended by a multi-hypothesis block pattern. This pattern indicates with one bit per 8 × 8 block whether one or two motion vectors and frame reference parameters are coded.

3. CODER CONTROL

In our coder control, a Lagrangian cost function is used for coding mode decisions. We adopt from [8] the relationship between the Lagrange parameter λ and the macroblock quantization parameter Q, given by

$$\lambda = 0.85Q^2. \tag{2}$$

Each coding mode decision incorporates the reconstruction error of the video signal as well as the bit-rate for each coding mode. The coding mode decisions are applied to all blocks, independent of their size.

On the macroblock level, we additionally have to decide between the INTER and INTER2H mode. For the INTER mode, we successively perform rate-constrained motion estimation (RC ME) for integer-pel positions and rate-constrained half-pel refinement. RC ME incorporates the prediction error of the video signal as well as the bit-rate for the motion vector and the picture reference parameter.

For the INTER2H mode, we perform rate-constrained multi-hypothesis motion estimation (RC MH ME). RC MH ME incorporates the multi-hypothesis prediction error of the video signal as well as the bit-rate for two motion vectors and picture reference parameters. RC MH ME is performed by the hypothesis selection algorithm, given in Fig. 2. This iterative algorithm performs conditional RC ME and is a low complexity solution to the joint estimation problem which has to be solved for finding an efficient pair of hypotheses $(\mathbf{c}_1, \mathbf{c}_2)^*$.

0: Assuming two hypotheses $\mathbf{c}_1$ and $\mathbf{c}_2$, the rate-distortion cost function

$$j(\mathbf{c}_1, \mathbf{c}_2) = \left\| \mathbf{s} - \frac{1}{2}\mathbf{c}_1 - \frac{1}{2}\mathbf{c}_2 \right\|_2^2 + \lambda\,[r(\mathbf{c}_1) + r(\mathbf{c}_2)]$$

is subject to minimization for each original block $\mathbf{s}$, given the Lagrange multiplier λ. Initialize the algorithm with two hypotheses $(\mathbf{c}_1^{(0)}, \mathbf{c}_2^{(0)})$ and set $i := 0$.

1: Minimize the rate-distortion cost function by full search for

a: hypothesis $\mathbf{c}_1^{(i+1)}$ while fixing hypothesis $\mathbf{c}_2^{(i)}$

$$\min_{\mathbf{c}_1^{(i+1)}} j(\mathbf{c}_1^{(i+1)}, \mathbf{c}_2^{(i)})$$

b: and hypothesis $\mathbf{c}_2^{(i+1)}$ while fixing the complementary hypothesis.

$$\min_{\mathbf{c}_2^{(i+1)}} j(\mathbf{c}_1^{(i+1)}, \mathbf{c}_2^{(i+1)})$$

2: As long as the rate-distortion cost function decreases, continue with step 1 and set $i := i + 1$.

Fig. 2. The proposed hypothesis selection algorithm is an iterative algorithm which successively improves two optimal conditional solutions.

Given the obtained motion vectors for the INTER and INTER2H modes, the resulting prediction errors are transform coded to compute the Lagrangian costs for the mode decision.

As already mentioned, multi-hypothesis motion-compensated prediction improves the prediction signal by spending more bits for the side-information associated with the motion-compensating predictor. But the encoding of the prediction error and its associated bit-rate also determines the quality of the reconstructed block. A joint optimization of multi-hypothesis motion estimation and prediction error encoding is far too demanding. But multi-hypothesis motion estimation independent of prediction error encoding is an efficient and practical solution. This solution is efficient if rate-constrained multi-hypothesis motion estimation, as explained before, is applied.

Testing the INTER4V mode, we apply the above method to 8 × 8 blocks. The Lagrangian costs of the four blocks as well as the

costs of the multi-hypothesis block pattern are added to compute the INTER4V costs.

It turns out that multi-hypothesis prediction is not the best mode for each block. The rate-distortion optimization therefore is a very useful tool to decide whether a block should be predicted with one or two hypotheses.

4. EXPERIMENTAL RESULTS

Our coder is based on the ITU-T Recommendation H.263 [4] and the results are comparable to those produced by the H.263 test model TMN-10 [9]. For our experiments the QCIF sequences *Foreman* and *Mobile & Calendar* are coded at 10 fps. Each sequence has a length of 10 seconds. We have investigated the influence of variable block-size (VBS) prediction and long-term memory (LTM) prediction on multi-hypothesis (MH) prediction.

Figs. 3 and 4 show the bit-rate values at 34 dB PSNR of the luminance signal over the number of reference frames M for the sequences *Foreman* and *Mobile & Calendar* respectively. We computed PSNR vs. bit-rate curves by varying the quantization parameter and interpolated intermediate points by a cubic spline. The performance of the codec with baseline prediction (BL), multi-hypothesis prediction (BL + MHP), variable block size prediction (BL + VBS), and multi-hypothesis prediction with variable block size (BL + VBS + MHP) is shown.

First, we investigate the influence of VBS prediction on MH prediction for one reference frame. VBS prediction is related to MH prediction in the way that more than one motion vector per macroblock is transmitted to the decoder. Both concepts, VBS as well as MH prediction provide gains for different scenarios. This can be verified by applying MH prediction to blocks of size 16×16 as well as 8×8. One bit for each block is sufficient to signal whether one or two prediction signals are used. To achieve a reconstruction quality of 34 dB in PSNR, the sequence *Mobile & Calendar* is coded in baseline mode with 389 kbit/s for $M = 1$. Correspondingly, MH prediction with $M = 1$ reduces the bit-rate to 351 kbit/s (See Fig. 4). We save about 10% of the bit-rate for MH prediction on macroblocks. Performing MH prediction additionally on 8×8 blocks, the rate of the bit stream is 332 kbit/s in contrast to 365 kbit/s for the codec with VBS. MH prediction saves about 9 % of the bit-rate produced by a codec with VBS prediction.

In summary, MH prediction works efficiently for both 16×16 and 8×8 blocks. The savings due to MH prediction are observed in the baseline mode as well as in the VBS prediction mode. Hence, our hypothesis selection algorithm in Fig. 2 is able to find two prediction signals in the previous frame which are combined more efficiently than just one prediction signal from the previous frame.

Second, we investigate the influence of long-term memory on MH prediction for variable block sizes. The multi-hypothesis codec with $M = 1$ reference frame has to choose both prediction signals from the previous frame. For $M > 1$, we allow more than one reference frame for each prediction signal. The reference frames for both hypotheses are selected by the rate-constrained multi-hypothesis motion estimation algorithm. The picture reference parameter allows also the special case that both hypotheses are chosen from the same reference frame. The rate constraint explained in the previous section is responsible for the trade-off between prediction quality and bit-rate. The performance of the MH codec with memory $M = 2, 5, 10$, and 20 is also depicted in Figs. 3 and 4. Going from one reference frame to 20 reference frames, the bit-rate is reduced from 332 to 247 kbit/s for the MH coder with variable block sizes when coding the sequence *Mobile & Calendar*. This corresponds to 25 % bit-rate savings. The long-term memory gain with VBS prediction is limited to 15 % for *Mobile & Calendar*. MH prediction benefits when being combined with long-term memory prediction so that the savings are more than additive. The bit-rate savings saturate for 20 reference frames for both sequences.

Figs. 5 and 6 depict the average luminance PSNR from reconstructed frames over the overall bit-rate produced by the codec with variable block size prediction (VBS) and with variable block size multi-hypothesis prediction (VBS+MHP) for the sequences *Foreman* and *Mobile & Calendar*. The number of reference frames is chosen to be $M = 1$ and $M = 20$.

We can also observe in these figures that MH prediction in combination with long-term memory compensation achieves coding gains up to 1.8 dB for *Foreman* and 2.8 dB for *Mobile & Calendar*. The reported coding gains correspond to bit-rate savings up to 23 % for *Foreman* and 32 % for *Mobile & Calendar*. It is also observed that the use of multiple reference frames enhances the efficiency of multi-hypothesis prediction.

5. CONCLUSIONS

The efficiency of rate-constrained multi-hypothesis prediction for video coding has been demonstrated. We focused on variable block size and long-term memory aspects for efficient video compression. We observed that VBS and MH prediction provide gains for different scenarios. MH prediction works efficiently for both 16×16 and 8×8 blocks. Also, it turns out that the use of long-term memory enhances the efficiency of multi-hypothesis prediction. The multi-hypothesis gain and the long-term memory gain do not only add up; MH prediction benefits from hypotheses which can be chosen from different reference frames. Multi-hypothesis prediction with long-term memory motion compensation achieves coding gains up to 2.8 dB, or equivalently, bit-rate savings up to 32 % for the sequence *Mobile & Calendar*. Therefore, multi-hypothesis prediction with long-term memory and variable block size turns out to be a very efficient concept for video compression.

6. REFERENCES

[1] G.J. Sullivan and R.L. Baker, "Rate-Distortion Optimized Motion Compensation for Video Compression Using Fixed or Variable Size Blocks," in *Proceedings of the IEEE Global Telecommunications Conference*, Phoenix, AZ, Dec. 1991, vol. 3, pp. 85–90.

[2] T. Wiegand, X. Zhang, and B. Girod, "Long-Term Memory Motion-Compensated Prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 70–84, Feb. 1999.

[3] 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.

[4] ITU-T, *Video Coding for Low Bitrate Communication: Recommendation H.263, Version 2*, 1998.

[5] M. Flierl, T. Wiegand, and B. Girod, "A Locally Optimal Design Algorithm for Block-Based Multi-Hypothesis Motion-

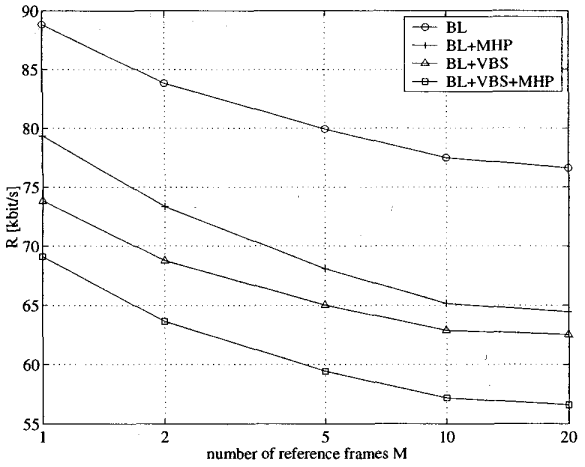

Fig. 3. Average bit-rate at 34 dB PSNR vs. number of reference frames for the QCIF sequence *Foreman.*

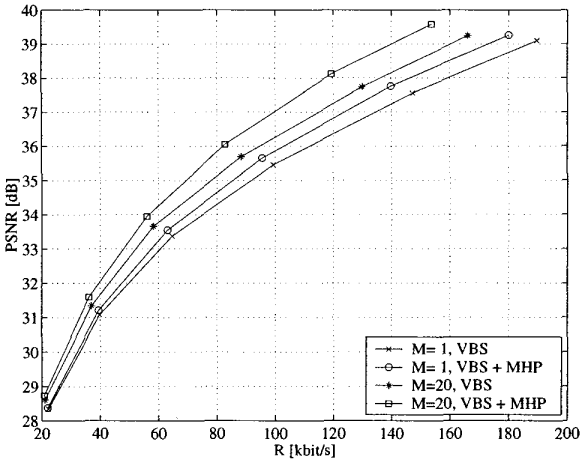

Fig. 5. PSNR of the luminance signal vs. overall bit-rate for the QCIF sequence *Foreman.*

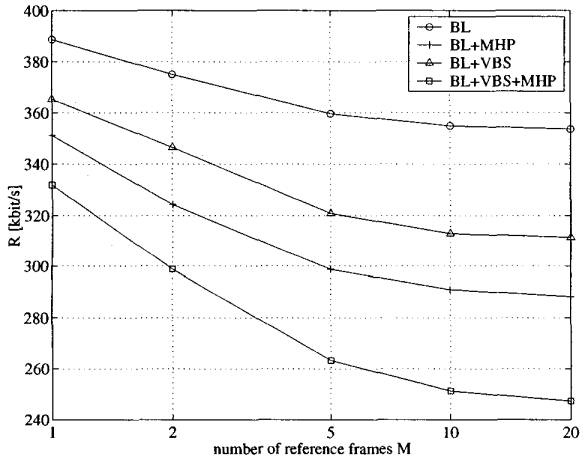

Fig. 4. Average bit-rate at 34 dB PSNR vs. number of reference frames for the QCIF sequence *Mobile & Calendar.*

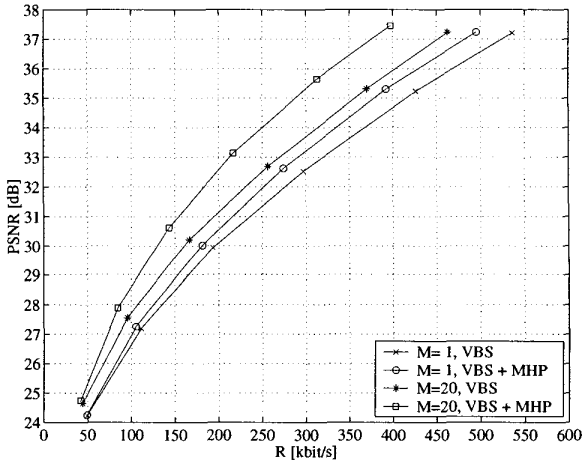

Fig. 6. PSNR of the luminance signal vs. overall bit-rate for the QCIF sequence *Mobile & Calendar.*

Compensated Prediction," in *Proceedings of the Data Compression Conference*, Snowbird, Utha, Apr. 1998, pp. 239–248.

[6] T. Wiegand, M. Flierl, and B. Girod, "Entropy-Constrained Linear Vector Prediction for Motion-Compensated Video Coding," in *Proceedings of the International Symposium on Information Theory*, Cambridge, MA, Aug. 1998, p. 409.

[7] M. Flierl, T. Wiegand, and B. Girod, "A Video Codec Incorporating Block-Based Multi-Hypothesis Motion-Compensated Prediction," in *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, Perth, Australia, June 2000.

[8] G.J. Sullivan and T. Wiegand, "Rate-Distortion Optimization for Video Compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 74–90, Nov. 1998.

[9] ITU-T/SG16/Q15-D-65, *Video Codec Test Model, Near Term, Version 10 (TMN-10), Draft 1*, Apr. 1998, Download via anonymous ftp to: standard.pictel.com/video-site/9804_Tam/q15d65d1.doc.