HIGH-ORDER MOTION COMPENSATION FOR LOW BIT-RATE VIDEO

Bernd Girod, Thomas Wiegand, Eckehard Steinbach, Markus Flierl, and Xiaozheng Zhang
Telecommunications Laboratory
University of Erlangen-Nuremberg
Cauerstrasse 7, 91058 Erlangen, Germany
e-mail: {girod,wiegand,steinb,flierl,zhang}@nt.e-technik.uni-erlangen.de

Invited Paper

ABSTRACT

The concept of high-order motion compensation is introduced. It is argued that in hybrid video coding, motion-compensated prediction has to be viewed as a source coding problem with a fidelity criterion. Based on our considerations, various high-order motion compensation approaches are presented that achieve significantly improved video coding results. The designed motion-compensated predictors achieve gains by exploiting high-order statistical dependencies in the video signal.

1 INTRODUCTION

In the early 1980s, video compression made the leap from intra-frame to inter-frame algorithms. Significantly lower bit-rates were achieved at the expense of memory and computational requirements that were two orders of magnitude larger. Today, with continuously dropping costs of semiconductors, we might soon be able to afford another leap by dramatically increasing the memory and computation power in video codecs. Algorithms to take advantage of increased memory and computation power are still in their infancy. This has been the motivation for our research into high-order motion-compensated prediction (MCP).

In recent years, several standards such as H.261, H.263, MPEG-1, and MPEG-2 have been introduced which mainly address the compression of video data for digital storage and communication services. H.263 [1] as well as the other standards utilize hybrid video coding schemes which consist of block-based MCP and DCT-based transform quantization of the prediction error. It is also highly likely that the future MPEG-4 standard [2] will follow the same video coding approach, but having a different application target as H.263.

2 BIT ALLOCATION IN HYBRID VIDEO CODING

The problem of optimum bit allocation to the motion vectors and the residual coding in any hybrid video coder is a non-separable problem requiring a high amount of computation. To circumvent this joint optimization, we split the problem into two parts: motion estimation and mode decision.

2.1 Rate-Constrained Motion Estimation

We can view MCP as a source coding problem with a fidelity criterion closely related to VQ. The rate-distortion trade-off can be controlled by various means. Our approach is to treat MCP as a special case of entropy-constrained vector quantization (ECVQ) [3].

The criterion for the block motion estimation is the minimization of a Lagrangian cost function

$$D_{DFD} + \lambda_{MOTION} R_{MOTION}, \tag{1}$$

in which the distortion D_{DFD}, representing the prediction error, is weighted against the rate R_{MOTION} associated with the motion parameters using a Lagrange multiplier λ_{MOTION}. The Lagrange multiplier imposes the rate constraint as in ECVQ, and its value directly controls the rate-distortion trade-off [3, 4, 5, 6].

2.2 Rate-Constrained Mode Decision

Hybrid video coding consists of the motion compensation and the residual coding stage. The task for residual coding is to represent signal parts that are not sufficiently compensated by motion coding. Hence, the distortion between the reconstructed and the original frame is either to be measured between the motion-compensated prediction signal or the signal after residual coding.

From the view-point of bit-allocation strategies, the various modes relate to various bit-rate partitions. Rate-constrained mode decision minimizes

$$D_{REC} + \lambda_{MODE} R_{TOTAL}, \tag{2}$$

for each macroblock. Again, the distortion after reconstruction D_{REC} is weighted against bit-rate using a Lagrange multiplier λ_{MODE}. But this time, the total bit-rate R_{TOTAL} to transmit a particular macroblock mode is considered, including the rates for the macroblock header, motion and texture coding. The coder can choose between various modes of operation. In particular, for H.263 [1], the coder can choose between inter-prediction modes INTER, INTER-4V which consist of motion compensation and residual coding, as well as the UNCODED mode which indicates copying the macroblock from the previous frame without residual coding, and the INTRA mode which is similar to JPEG image coding.

3 LONG-TERM MEMORY MOTION COMPENSATION

Long-term memory MCP extends the motion vector utilized in hybrid video coding by a variable time delay permitting the use of several decoded frames instead of only the previously decoded one for block-based motion compensation. In this paper, we restrict the maximum number of frames in the long-term memory to 50 corresponding to decoded video frames of 5 seconds at 10 frames/s sampling rate. The frames inside the long-term memory which is simultaneously built at encoder and decoder are addressed by a combination of the codes for the spatial displacement vector and the variable time delay. Hence, the transmission of the variable time delay potentially increases the bit-rate which has to be justified by improved MCP.

This trade-off limits the efficiency of the proposed approach. Therefore, we control the motion bit-rate by employing rate-constrained motion estimation where a Lagrangian cost function as given in (1) is minimized. Figure 1 shows the result of a motion compensation experiment where the rate constraint is employed. The plots show the motion-compensated prediction error measured as PSNR in dB vs. bit-rate for the motion vectors measured over 100 frames. We predict all frames between 200-299 while the frame skip parameter relates to the sub-sampling of the reference frames which are the original frames in this experiment. The motion search is conducted by full search in the range $M \times [-15...15] \times [-15...15]$ on integer-pel positions followed by half-pel refinement. As criterion for the block motion search we use the sum of the squared differences (SSD) between displaced and original frame (D_{DFD}) while the Lagrange parameter (λ_{MOTION}) is varied over values 0, 1, 2, 5, 10, 20, 50, 100, 500, 1000, 2000, 5000, and 10000. For the motion vectors, the H.263 MVD table is used employing the H.263 motion vector median prediction [1]. Four cases are shown:

1. Motion compensation using the last frame only while allowing only blocks of size 16×16
2. As in 1, but also blocks of size 8×8 are permitted
3. As in 1, but the last $M = 10$ frames are used for motion compensation
4. Combination of cases 2 and 3

Let us first consider cases 1 and 2. These two cases are distinguished by allowing the use of blocks of size 8×8 in addition to the 16×16 blocks for motion compensation. The PSNR values achievable using 8×8 blocks are much higher. However, also the bit-rate that can possibly be spent drastically increases. Hence, bit allocation is very important for case 2 while for case 1, the coder could also be operated at $\lambda_{MOTION} = 0$ without much loss in performance. Bit allocation becomes even more important when we permit 10 frames instead of 1 frame for motion-compensated prediction. Finally, case 4, i.e., the combination of using the long-term memory and 16×16 as well as 8×8 blocks provides the largest gains.

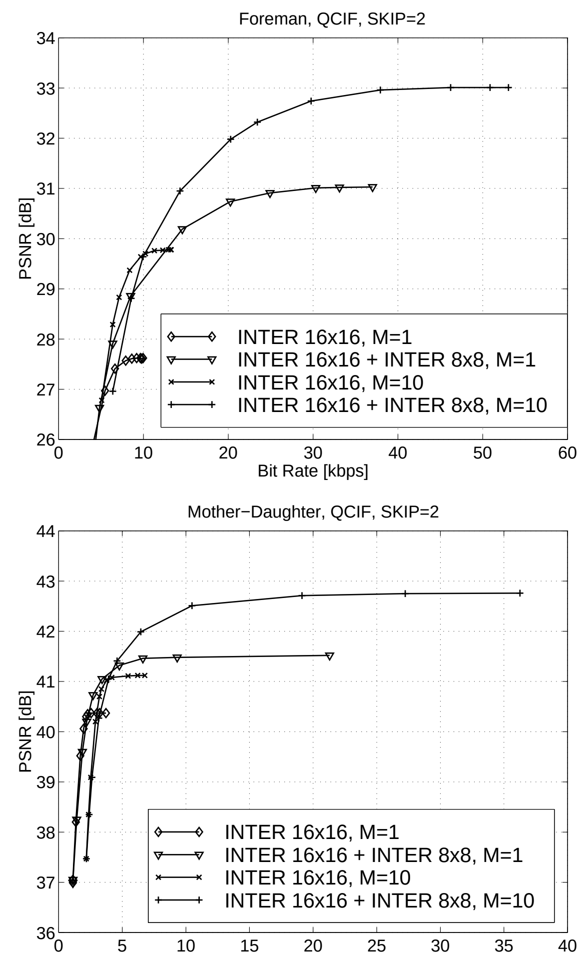

Figure 1: Long-term memory prediction: prediction error measured in PSNR vs. bit-rate for motion parameters for the sequences *Foreman* (top) and *Mother-Daughter* (bottom).

Simulation results are obtained by integrating long-term memory prediction into an H.263 codec. For that, case 4 in Fig. 1 is employed for MCP. Also rate-constrained mode decision is used as described above. The result is presented in Fig. 2. The baseline for our comparisons is the TMN-2.0 codec, a public available software H.263 encoder [7] (labeled as TMN in Fig. 2). The TMN-2.0 encoding scheme does not employ rate-constrained motion estimation and uses heuristic criteria for the mode decision. In contrast, our rate-distortion optimized H.263 coder uses rate-constrained motion estimation as well as rate-constrained mode decision (labeled as H.263 in Fig. 2) On top of that, the long-term memory coder is run employing the same encoding strategy as the rate-distortion optimized H.263 coder. The long-term memory is varied over M =2, 5, 10, and 50.

Reconstruction PSNR improvements up to 2 dB for the *Foreman* sequence and 1.5 dB for the *Mother-Daughter* sequence are demonstrated in comparison to the TMN-2.0 H.263 coder as shown in Fig. 2. The PSNR improvements correspond to bit-rate savings up to 34 % and 30 %, respectively. These bit-rate savings split into 11 % (*Foreman*) and 13 % (*Mother-Daughter*) due to our modifications to the encoding strategy and into 23 % (*Foreman*) and 17 % (*Mother-Daughter*) due to the impact of long-term memory MCP. For more details on long-term memory MCP, please refer to [8].

4 MULTI-HYPOTHESIS MOTION-COMPENSATION

Multi-hypothesis MCP extends the prediction model to a generalized approach where various video signals are combined using linear superposition [9, 10]. The approach is similar to sub-pel accurate MCP or B-frames, however, it allows arbitrary combinations of blocks that are addressed using the long-term memory prediction scheme.

The multi-hypothesis motion-compensated predictor is defined as

$$\hat{\mathbf{S}} = \sum_{n=1}^{N} h_n \cdot \mathbf{C}_n \,, \qquad (3)$$

with $\hat{\mathbf{S}}$ and $\mathbf{C}_n$ being K-tuples with K typically being 256 for 16×16 blocks or 64 for 8×8 blocks. This scheme is a generalization of sub-pel accurate MCP and B-frames. In order to incorporate spatial filtering and overlapped block motion compensation (OBMC), the scalars h_n have to be matrices of dimension $K \times K$ and the OBMC window is represented by elements of this matrix which are on the main diagonal.

In our experiments, the superposition coefficients h_n are fixed to the value $1/N$, but we conduct a search to

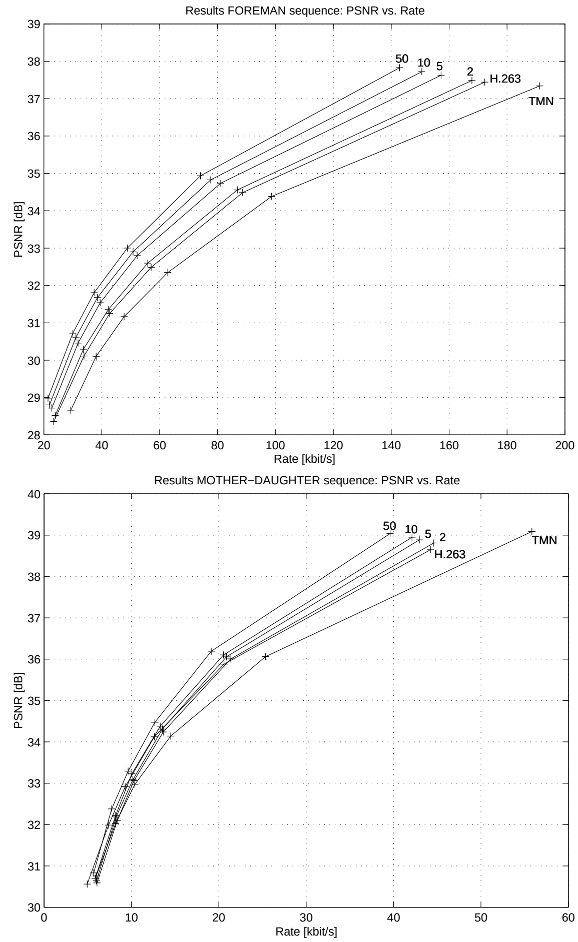

Figure 2: PSNR vs. overall bit-rate for the sequences *Foreman* (top) and *Mother-Daughter* (bottom). The quantization parameter is varied over the values 7, 10, 3, 16, 22, and 31.

find the optimum input vectors, which are mutually dependent. Hence, in general, the entire product space of the N hypotheses has to be searched. An iterative algorithm avoids searching the complete space by successively improving n optimal conditional solutions [9, 10]. Convergence to a local optimum is guaranteed, because the algorithm prohibits an increase of the error measure.

We control the rate of the MC data that have to be transmitted as side information by minimizing the Lagrangian cost function in (1). An adaptive algorithm for optimally selecting the number of input blocks N is given in [9, 10].

Figure 3 shows the result of a motion compensation experiment involving multi-hypothesis motion compensated prediction. The simulation conditions are the same as for Fig. 1. Four cases are shown:

1. Motion compensation using the 10 frames while allowing only blocks of size 16×16 and only N=1

hypothesis

2. As in 1, but also blocks of size 8×8 are permitted
3. As in 1, but multi-hypothesis prediction is used permitting up to N=4 hypotheses to be superimposed for a 16×16 block
4. Combination of cases 2 and 3

Cases 1 and 2 are the same as cases 3 and 4 in Fig. 1. Case 3 performs worse than case 2 for *Foreman* and better for *Mother-Daughter*. Finally, for case 4 that is combining cases 2 and 3, we always obtain superior rate-distortion performance. Note, that in contrast to 8×8 block coding where the H.263 motion vector prediction is used, the multi-hypothesis motion coder does not employ motion vector prediction. Hence, there is a potential direction for improvement of multi-hypothesis MCP in terms of coding efficiency.

Adaptive block size and multi-hypothesis MCP are not necessarily competing concepts for motion video coding. Also 8×8 blocks can be coded using a multi-hypothesis scheme. Finally, the question arises about the gains of multi-hypothesis coding when incorporated into a hybrid video codec. This item is subject of ongoing research.

5 CONCLUDING REMARK

High-order motion compensated prediction is a straightforward extension of current state-of-the-art low bit-rate video coding schemes. Our experimental results that have included up to 50 previous frames are very encouraging. We expect to see compression algorithms based on many past frames to become wide-spread in the future. The increase in memory associated with the approach should be insignificant soon.

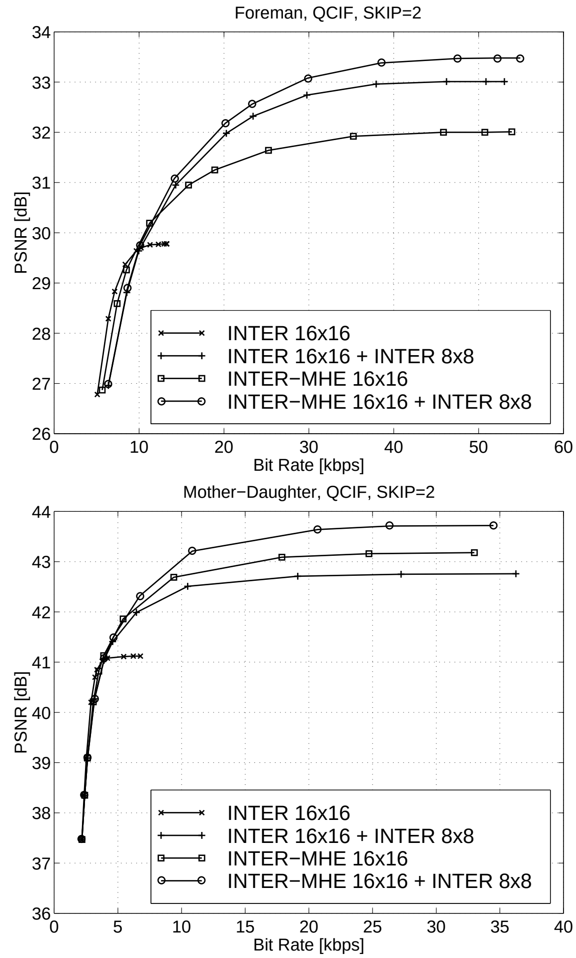

Figure 3: Multi-hypothesis prediction: prediction error measured in PSNR vs. bit-rate for motion parameters for the sequences *Foreman* (top) and *Mother-Daughter* (bottom).

References

[1] ITU-T Recommendation H.263, "Video Coding for Low Bitrate Communication", June 1996.

[2] ISO/IEC JTC1/SC29/WG11, "MPEG-4 Video Verification Model", Draft, July 1997.

[3] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-Constrained Vector Quantization", *IEEE Trans. ASSP*, vol. 37, no. 1, pp. 31–42, Jan. 1989.

[4] Y. Shoham and A. Gersho, "Efficient Bit Allocation for an Arbitrary Set of Quantizers", *IEEE Trans. ASSP*, vol. 36, pp. 1445–1453, Sept. 1988.

[5] G. J. Sullivan and R. L. Baker, "Rate-Distortion Optimized Motion Compensation for Video Compression Using Fixed or Variable Size Blocks", in *GLOBECOM'91*, 1991, pp. 85–90.

[6] B. Girod, "Rate-Constrained Motion Estimation", in *Proc. SPIE VCIP*, Chicago, USA, Sept. 1994, pp. 1026–1034, (Invited paper).

[7] Telenor Research, "TMN (H.263) Encoder/Decoder, Version 2.0", Download via anonymous ftp to `bonde.nta.no`, June 1997.

[8] T. Wiegand, X. Zhang, and B. Girod, "Long-Term Memory Motion-Compensated Prediction", *IEEE Trans. CSVT*, Sept. 1998, To appear. Download: http://www-nt.e-technik.uni-erlangen.de/ ˜ wiegand/trcsvt98{.ps.gz—.pdf}.

[9] M. Flierl, T. Wiegand, and B. Girod, "A Local Optimal Design Algorithm for Block-Based Multi-Hypothesis Motion-Compensated Prediction", in *Proc. DCC*, Snowbird, USA, Mar. 1998.

[10] T. Wiegand, M. Flierl, and B. Girod, "Entropy-Constrained Linear Vector Prediction for Motion-Compensated Video Coding", in *Proc. ISIT*, Boston, USA, Aug. 1998, Submitted for publication.