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

ABSTRACT

A generalized concept for motion-compensated prediction that covers approaches like sub-pel motion compensation, overlapped block motion compensation and B-frames is presented. Reflections on the determination and encoding of the motion parameters involving image segmentation, motion models, and motion parameter prediction are presented. It is suggested 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 referenced that achieve significantly improved video coding results. The designed motion-compensated predictors achieve gains by exploiting high-order redundancies in the video signal.

1 INTRODUCTION

Motion-compensated prediction (MCP) schemes achieve data compression by exploiting correlation in video signals. Often, with such schemes, MCP is combined with intraframe encoding of the prediction error. Successful applications range from digital video broadcasting to low bit-rate video phones.

Utilizing inter-frame coding for video compression leads to the question of the rate-distortion efficiency of MCP, since the bit-rate for the motion parameters has to be taken into account. Thus, for a certain bit-rate required to transmit the motion-related information, MCP provides a version of the video signal with a certain distortion.

In section 2, we give an overview of existing motion compensation (MC) methods, where various known approaches to MCP are viewed regarding their rate-distortion and computational efficiency. Then, in section 3, the descriptive aspect of MCP is emphasized, viewing it as a source coding problem with a fidelity criterion. In section 4, several design ideas for motion-compensated predictors are presented that take advantage of the considerations in sections 2 and 3.

2 MOTION-COMPENSATED PREDICTION IN VIDEO COMPRESSION

Motion compensation can be divided into two parts: the intensity prediction model to produce the samples of MCP signal and the encoding of the motion model parameters for efficient transmission of the intensity predictor field.

2.1 Intensity Prediction Model

In the video coding literature and ITU as well as ISO video coding standards, several approaches to MCP have been proposed including

1. copying samples from previously reconstructed frames [1],
2. sub-pel accurate MC [2, 3],
3. multi-hypothesis MCP, including overlapped block MC (OBMC) [4] and B-frames [5],

Item 1, proposed in the early 1980s by Jain and Jain [1], refers to copying samples from the previously decoded frame. It was the first step that made the leap from intra-frame to inter-frame algorithms improving rate-distortion performance of video coding. Significantly lower bit-rates were obtained by exploiting low-order redundancies in the video signal. These gains were achieved at the expense of memory and computational requirements that were two orders of magnitude larger. Since then, more complex prediction models have appeared: sub-pel accurate MC [2, 3] and multi-hypothesis MCP including overlapped block motion compensation (OBMC) [4] and B-frames [5].

Let us assume, the MCP signal $\hat{s}[x, y]$ at sampling position x, y to be given as

$$\hat{s}[x, y] = \mathbf{f}^T[x, y] \cdot \mathbf{c}(\mathbf{p}[x, y]), \qquad (1)$$

where $\mathbf{c}$ is a vector valued signal containing samples from previously decoded frames that are addressed by $\mathbf{p}[x, y]$, which are the motion parameters for sampling position

x, y. The filter $\mathbf{f}[x, y]$ in general also depends on the sampling position x, y.

For item 2, sub-pel accurate MC, the samples in $\mathbf{c}$ come from adjacent integer-pel positions of previously decoded frames, while $\mathbf{p}[x, y]$ indexes them separately or points to the exact sub-pel position using just one index. An example for an efficient implementation is the specification of half-pel accurate MC in H.263 [6]. OBMC and B-frames modify the condition for sub-pel accurate MC, namely that the samples in $\mathbf{c}$ have to come from adjacent integer-pel positions.

For item 3, OBMC, several samples are indexed by various motion parameters that are assigned spatially adjacent blocks and the linear superposition using $\mathbf{f}[x, y]$ is performed dependent on the position of the sample relative to the block centers. The vector $\mathbf{f}[x, y]$ contains the OBMC weights that are dependent on x and y. An example for an implementation of OBMC is given in Annex F of the ITU-T recommendation H.263 [6] where the concepts of half-pel accurate MC and OBMC are combined.

For the other technique of item 3, B-frames, the samples in $\mathbf{c}$ come from temporally ensuing and preceding frames. In most cases, $\mathbf{p}$ indexes the two samples that are superimposed by two separate motion parameters. B-frames and half-pel accurate MC have been successfully combined in Annex O of the ITU-T recommendation H.263 [6]. Note that the complexity increase compared to copying samples is considerable. In general, all combinations allowed for sub-pel accurate MC, OBMC or B-frames has to be searched to achieve optimum performance.

The theoretical gains achievable by superposition of various signal for MCP are analyzed in [7].

2.2 Encoding of the Motion Model Parameters

In order to control the bit-rate for motion model parameters, the linear intensity predictors for each sample (sub-pel accurate and multi-hypothesis MCP) are quantized. The methods employed for this task are mostly related to regression methods, since early ideas on the subject emphasized the aspect of denoising a "true motion" field. In principle, every method that takes advantage of the memory in the linear sample predictor field could be useful. However, the performance of a method for quantization of the linear sample predictor field to lower the bit-rate for motion parameters has to be evaluated including measures on its rate-distortion and computational efficiency. Several aspects are involved including

1. image segmentation,
2. motion models,
3. motion parameter prediction and entropy coding.

Item 1, image segmentation, is a means for indicating the use of various quantizers/motion-compensated predictors that have been adaptively chosen by the motion estimator for the corresponding image segments. The image segmentation using blocks of fixed and variable size [8] or regions of arbitrary size can be viewed together, in that region-based coding can be seen as an extension of variable-block-size MC to arbitrary shapes. The usefulness of fixed and variable block size MC is well understood, and led to its incorporation into the ITU-T Recommendation H.263 [6]. In variable block size MC, in most cases, the image partition is obtained by recursive splitting of image segments. If only splitting is allowed, the motion search in most cases consists of minimizing an affine tree-functional mapped on the variable block-size partition that can be done optimally by tree pruning [9]. In contrast, the difficulties in demonstrating the efficiency of region-based coding mainly relate to the mutual dependencies that are introduced by the selection of the regions, their contours, and coding parameters and the bit-rate associated with signaling them.

Item 2, the motion model, is highly connected to image segmentation. The methods employed include translational motion, and complex motion models such as affine (6 parameter) or bilinear (12 parameter) polynomial motion models [10]. Translational motion can be viewed as subset of the complex motion models where only the constant horizontal and vertical shifts are unequal to zero. The motivation for increasing the complexity (bit-rate) of the motion model lies in the fact that large regions of images in video sequences are not likely to be motion-compensated by a constant linear predictor (translational motion).

When covering a large segment of the image by a complex motion model, e.g. an affine motion model, we assume that all samples of the segment can be sufficiently motion-compensated by sample-based linear predictors which parameters are encoded using the motion model. Another concept is given by item 3, the prediction of the codes assigned to fractions of the large segment. Most of the time, these codes are predicted using data already available at the decoder such as codes assigned to previously encoded parts of the image. In terms of rate-distortion efficiency this method can never be superior to joint coding of the fractions of the image segments. However, it reveals the benefit of low complexity for encoding the smaller fractions of the large segment. The prediction of motion parameters is mainly emphasized in the context of block-based translational motion compensation.

3 MOTION-COMPENSATED PREDICTION AS A SOURCE CODING PROBLEM

In terms of rate-distortion efficiency, we can view motion-compensated prediction as follows: *Motion-compensated prediction is the quantized prediction of a multi-dimensional random variable by quantized realizations of itself.* The term *quantized prediction* refers to the fact that in a hybrid video coding environment, the motion parameters have to be transmitted as side information. The term *quantized realizations of itself* relates to the assumption of memory inside the *multi-dimensional random variable* and that MC is performed simultaneously at the encoder and decoder using transmitted data.

This definition is very close to that of vector quantization (VQ). Hence, we can view MCP as a source coding problem with a fidelity criterion being highly related to VQ. For a certain bit-rate required to transmit the motion parameters, MCP provides a version of the video signal with a certain distortion. 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) [11]. The image blocks to be encoded are quantized using their own code books that consist of image blocks of the same size in frames that are available at encoder and decoder. A code book entry is addressed by the motion parameters, which are entropy-coded. The criterion for the block motion estimation is the minimization of a Lagrangian cost function

$$D + \lambda R, \tag{2}$$

in which the distortion D represented by the prediction error, is weighted against the rate R associated with the motion parameters using a Lagrange multiplier λ. The Lagrange multiplier imposes the rate constraint as in ECVQ, and its value directly controls the rate-distortion trade-off [11, 12, 13, 14].

When increasing the capability of MCP to provide versions of the video signal at higher fidelity we also increase the number of bits that can possibly be spent on the motion parameters. For example, reducing the block size down to only one pixel per block may lead to a perfect description of the video signal, however it will produce a high bit-rate too. Therefore, when the description achievable by MCP is adjustable over a wide range of fidelities and bit-rates, which is the case for high-order MCP, bit allocation methods become essential.

4 HIGH-ORDER MOTION-COMPENSATED PREDICTION

Summarizing the previous two sections, we can draw the following conclusions:

- Sub-pel accurate MCP, OBMC, and B-frames correspond to combining several signals by linear superposition for motion-compensated prediction, allowing MCP with higher fidelity at the cost of increased bit-rate and computational complexity.
- Methods for encoding of the motion parameters involving image segmentation, motion models, and motion parameter prediction take advantage of memory in the predictor field, also allowing us to trade-off fidelity of the description, bit-rate and complexity.

In what follows, we have tried to emphasize these aspects in various design approaches.

The first restriction we have relaxed is the use of only the previously decoded frame for MCP. The approach is called long-term memory MCP, where the spatial displacement vector used in block-based hybrid video coding is extended by a variable time delay permitting the use of more frames than the previously decoded one for MCP. The long-term memory typically covers several seconds of decoded frames at encoder and decoder. The use of multiple frames for MC in most cases provides significantly improved prediction gain. The variable time delay has to be transmitted as side information requiring additional bit-rate, which may be prohibitive when the size of the long-term memory becomes too large. Therefore, we control the bit-rate of the motion information by employing rate-constrained motion estimation where a Lagrangian cost function as given in section (2) is minimized. Simulation results are obtained by integrating long-term memory prediction into an H.263 codec. 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. The PSNR improvements correspond to bit-rate savings up to 34 % and 30 %, respectively. Mathematical inequalities are used to speed-up motion estimation while achieving full prediction gain. For details, please refer to [15].

In [16], the restriction to use previously decoded frames for MCP has been modified to use warped versions of the previously decoded frame. This scheme is similar to global MC. In contrast to global MC, where typically one motion model is transmitted, we show that in the general case more than one motion model is of benefit in terms of coding efficiency. The various reference frames are addressed as in long-term memory MCP. Relating the approach to region-based coding with polynomial motion models, we note that we also transmit various motion parameter sets and that the various "regions" associated with these motion parameter sets are indicated by the frame selection parameter. But, the "regions" in our scheme do not have to be connected. They

are restricted to the granularity of the fixed or variable block-size segmentation of the block-based video codec. Furthermore, each block belonging to a "region" may have an individual spatial displacement vector. This is beneficial if the motion exhibited in the scene cannot be compensated by a few polynomial motion models. In addition, if the video scene does not lend itself to a description by various polynomial motion models, the coder drops into its fall-back mode, block-based MC using the previously decoded frame only. The approach is also incorporated into an H.263-based video codec and embedded into a rate-constrained motion estimation and macroblock mode decision frame work. PSNR gains of 1.2 dB in comparison to the H.263 codec for the high global and local motion sequence *Stefan* and 1 dB for the sequence *Mobile & Calendar*, which contains no global motion, are reported. These PSNR gains correspond to bit-rate savings of 21 % and 30 % compared to the H.263 codec, respectively [16].

Finally, we extended the prediction model to a generalized approach where various video signals are combined using linear superposition [17, 18]. The approach is very similar to sub-pel accurate MCP or B-frames, however, allows arbitrary combinations of blocks that are addressed using the long-term memory prediction scheme. The superposition coefficients are fixed, but we conduct a search to find the optimum input vectors, which are mutually dependent. We control the rate of the MC data that have to be transmitted as side information by minimizing the Lagrangian cost function in Eq. (2). An adaptive algorithm for optimally selecting the number of input blocks is given. The designed motion-compensated predictors show PSNR gains in prediction error up to 4.4 dB at the cost of increased bit-rate of 16 kbit/s when comparing them to conventional MCP for the sequence *Foreman*.

5 FINAL REMARKS AND FUTURE WORK

Finally, the question arises about the limits of the coding gains achievable by high-order MCP. Also, the question to what extend the added complexity will justify the coding gains has to be discussed. Hence, future work is concentrated on further improving the presented schemes and combining them in one video codec. Also, the reduction of complexity remains an important issue to be addressed in order to make the methods more suitable for practical implementations.

References

[1] J. R. Jain and A. K. Jain, "Displacement Measurement and Its Application in Interframe Image Coding", *IEEE Trans. COM*, vol. 29, no. 12, pp. 1799–1808, Dec. 1981.

[2] B. Girod, "The Efficiency of Motion-Compensating Prediction for Hybrid Coding of Video Sequences", *IEEE JSAC*, vol. 5, no. 7, pp. 1140–1154, Aug. 1987.

[3] B. Girod, "Motion-Compensating Prediction with Fractional-Pel Accuracy", *IEEE Trans. COM*, vol. 41, no. 4, pp. 604–612, Apr. 1993.

[4] H. Watanabe and S. Singhal, "Windowed Motion Compensation", in *Proc. SPIE VCIP*, 1991, vol. 1605, pp. 582–589.

[5] ISO/IEC 13818 (ITU-T H.262), "Generic Coding of Moving Pictures and Associated Audio Information, Part 2: Video", International Standard, Mar. 1994.

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

[7] B. Girod, "Efficiency Analysis of Multi-Hypothesis Motion-Compensated Prediction for Video Coding", *IEEE Trans. IP*, 1997, Submitted for publication.

[8] G. J. Sullivan and R. L. Baker, "Efficient Quadtree Coding of Images and Video", *IEEE Trans. IP*, vol. 3, no. 3, pp. 327–331, May 1994.

[9] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Optimal Pruning with Applications to Tree-Structured Source Coding and Modeling", *IEEE Trans. IT*, vol. 35, no. 2, pp. 299–315, Mar. 1989.

[10] ISO/IEC JTC1/SC29/WG11 MPEG96/M0904, "Nokia research center: Proposal for efficient coding", Submitted to Video Subgroup, July 1996.

[11] 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.

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

[13] 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.

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

[15] 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}.

[16] T. Wiegand, E. Steinbach, A. Stensrud, and B. Girod, "Multiple Reference Picture Coding using Polynomial Motion Models", in *Proc. SPIE VCIP*, San Jose, USA, Feb. 1998.

[17] 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.

[18] 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.