MULTIHYPOTHESIS PICTURES FOR H.26L

Markus Flierl
Telecommunications Laboratory
University of Erlangen-Nuremberg
Erlangen, Germany
mflierl@stanford.edu

Thomas Wiegand
Image Processing Department
Heinrich Hertz Institute
Berlin, Germany
wiegand@hhi.de

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

ABSTRACT

Multihypothesis pictures are related to B pictures where two motion-compensated signals from previous and future reference frames are linearly combined for prediction. Multihypothesis pictures also linearly combine two motion-compensated signals but only from temporally previous reference frames. Consequently, multihypothesis pictures have the advantage that there is no extra coding delay like for B pictures. In addition, multihypothesis pictures are also used for reference to predict future multihypothesis pictures. We integrate multihypothesis prediction into the emerging H.26L standard by introducing one additional multihypothesis macroblock mode for low delay video coding which is similar to the bi-directional macroblock mode for B pictures. The current implementation achieves bit-rate savings up to 13% for the investigated test-sequences when compared to the H.26L test model.

1. INTRODUCTION

Significant coding gains have been reported when adding multihypothesis motion-compensated prediction in combination with multiple reference pictures to the ITU-T standard H.263 [1, 2]. The efficiency analysis in [3] provides theoretical insight about multihypothesis motion-compensated prediction. An extension to optimal multihypothesis motion estimation can be found in [4]. The relative gain by multihypothesis prediction can be improved by increasing the number of available reference pictures at encoder and decoder as reported in [2]. This observation is supported by analytical investigations on multihypothesis motion-compensated prediction with forward-adaptive hypothesis switching in [5].

The emerging ITU-T standard H.26L [6] supports motion compensation with 7 different block sizes and quarter pel accuracy and utilizes a 4×4 integer transform for residual coding. This paper investigates multihypothesis pictures for the ITU-T standard H.26L. Multihypothesis pictures utilize multihypothesis motion-compensated prediction. Two motion-compensated macroblocks from temporally previous pictures are linearly combined to improve the efficiency of motion-compensated prediction where both macroblocks have individual block types (sub-block sizes) and picture reference parameters.

This paper is organized as follows: In section 2, we explain multihypothesis pictures for H.26L and introduce an additional multihypothesis macroblock mode. Section 3 discusses encoder issues like rate-constrained multihypothesis motion estimation and coding mode decision. Section 4 presents experimental results for multihypothesis pictures and compares them to the H.26L P pictures.

Markus Flierl currently visits the Information Systems Laboratory at Stanford University.

2. MULTIHYPOTHESIS PICTURES

Multihypothesis motion pictures [7] are an extension of P pictures such that each macroblock can be compensated by a linear combination of two motion-compensated macroblocks. Conventional B pictures also employ two linearly combined motion-compensated blocks but one motion-compensated signal (hypothesis) origins from a future reference frame. In contrast to B pictures, multihypothesis motion pictures (MH pictures) utilize temporally previous pictures for prediction and cause no extra coding delay. In addition, decoded MH pictures are also used for reference to predict future MH pictures.

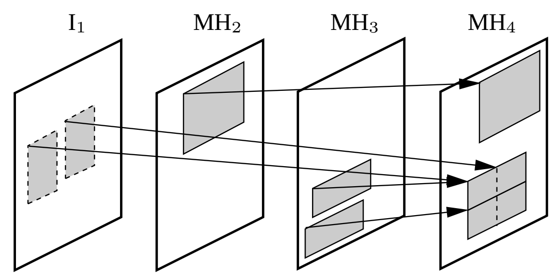

Fig. 1. Multihypothesis motion pictures utilize long-term memory motion-compensated prediction and multihypothesis motion-compensated prediction. For multihypothesis prediction, just two macroblock-hypotheses (or macrohypotheses) from temporally previous pictures are used to predict the current macroblock of the MH picture.

Fig. 1 visualizes multihypothesis pictures. They include conventional long-term memory motion-compensated prediction. In addition, they also allow linearly combined motion-compensated macroblocks with individual block size types (macrohypotheses). Both macrohypotheses are just averaged to form the current macroblock. For macroblocks, the H.26L test model [6] allows 7 different block sizes that range from 16×16 to 4×4. For multihypothesis prediction, we utilize 7 macrohypothesis types. MH pictures use like P pictures temporally previous pictures for prediction.

For efficient long-term memory motion-compensated prediction, H.263 utilizes picture reference parameters for macroblocks and 8×8 blocks [8]. The current H.26L test model just allows one picture reference parameter for the macroblock and assumes that all sub-blocks can be found on that specified reference picture. Experimental results support this design: When modifying the test model such that an individual picture reference parameter is used for each block, almost identical compression efficiency is observed for high bit-rates. For low bit-rates, however, the compression efficiency may decrease due to the larger overhead introduced by transmitting picture reference parameters for each block. Fig. 2 compares the H.26L test model TML-6 to this modified codec.

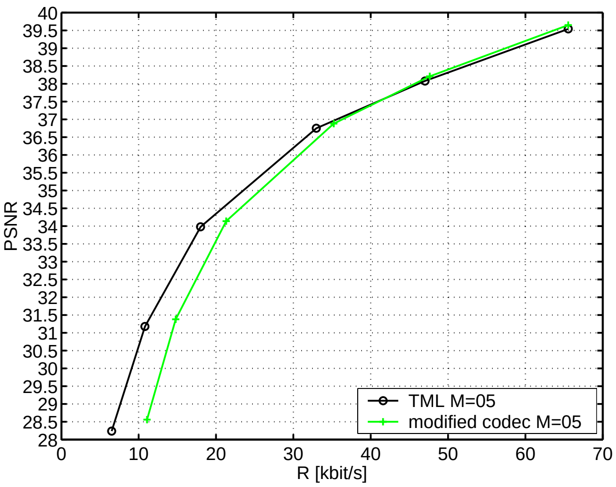

Fig. 2. PSNR of the luminance signal vs. overall bit-rate for the QCIF sequence *Container* with 10 fps and 5 reference frames.

On the other hand, the individual picture reference parameter per block is efficient for H.263-based multihypothesis motion-compensated prediction [2]. Previous work on this topic investigates multihypothesis motion estimation for arbitrarily sized blocks with individual picture reference parameters, that is, one block is associated with one hypothesis. But with respect to the H.26L design, one hypothesis has to be associated with one macroblock, which can be composed of smaller blocks that originate from the same reference picture. In the following, we discuss multihypothesis motion-compensated prediction with macrohypotheses.

2.1. Structure of the Multihypothesis Mode

Macroblocks in the current frame predicted by the multihypothesis mode result from a linear combination of two macrohypotheses. Each macrohypothesis is selected from 7 macrohypothesis types that correspond to 7 block size types specified in the H.26L test model for bi-directional prediction as depicted in Fig. 3.

In the multihypothesis mode, two macrohypotheses are signaled. Each macrohypothesis is specified by a picture reference parameter, a macrohypothesis type, and the corresponding motion vector for each sub-block.

3. ENCODER

In our coder control, a Lagrangian cost function is used for coding mode decisions [9]. For the multihypothesis mode, we utilize rate-constrained multihypothesis motion estimation. The cost function

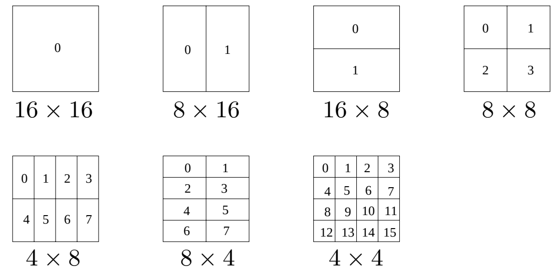

Fig. 3. Macrohypothesis types for the multihypothesis mode.

incorporates the multihypothesis prediction error of the video signal as well as the bit-rate for two picture reference parameters, two macrohypotheses types, and the associated motion vectors. Rate-constrained multihypothesis motion estimation is performed by the macrohypothesis selection algorithm. This iterative algorithm performs conditional rate-constrained motion estimation and is a low complexity solution to the joint estimation problem which has to be solved for finding an efficient pair of macrohypotheses.

The iterative algorithm is initialized with the data of the best macroblock type for long-term memory motion-compensated prediction (initial macrohypothesis). The algorithm continues with:

1. One macrohypothesis is fixed and conditional rate-constrained motion estimation is applied to the complementary macrohypothesis such that the multihypothesis costs are minimized.
2. The complementary macrohypothesis is fixed and the first macrohypothesis is optimized.

The two steps are repeated until convergence. For the current macrohypothesis, conditional rate-constrained motion estimation determines the conditional optimal picture reference parameter, macrohypothesis type, and associated motion vectors. For the conditional motion vectors, an integer-pel accurate estimate is refined to half-pel and quarter-pel accuracy.

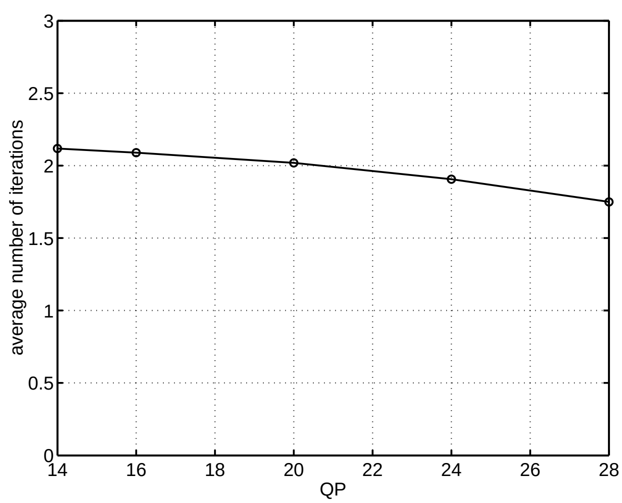

Fig. 4. Average number of iterations for multihypothesis motion estimation vs. quantization parameter for the CIF sequence *Mobile & Calendar* with 30 fps and $M = 5$ reference frames.

Fig. 4 shows the average number of iterations for multihypothesis motion estimation with 5 reference frames over the quantization parameter. It takes about 2 iterations to achieve an error smaller than 0.5% relative to the error in the previous iteration. The algorithm converges faster for higher quantization parameter values.

Given the best single macrohypothesis for motion-compensated prediction and the best macrohypothesis pair for multihypothesis prediction, the resulting prediction errors are transform coded to compute the Lagrangian costs for the mode decision.

Multihypothesis 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 macroblock. A joint optimization of multihypothesis motion estimation and prediction error coding is far too demanding. But multihypothesis motion estimation independent of prediction error encoding is an efficient and practical solution if rate-constrained multihypothesis motion estimation is applied.

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

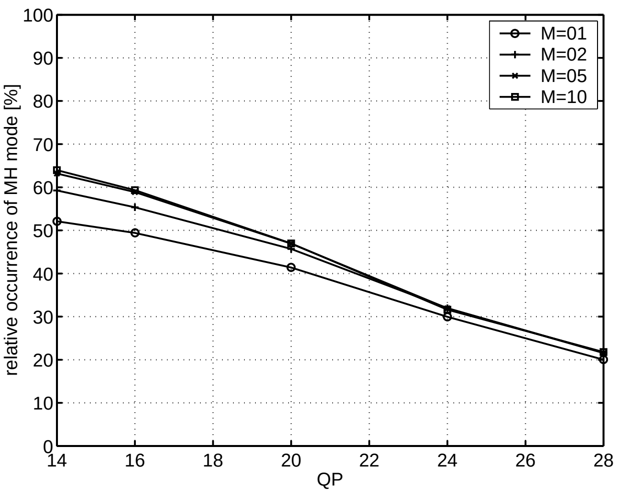

Fig. 5. Relative occurrence of the multihypothesis mode vs. quantization parameter for the CIF sequence *Mobile & Calendar* with 30 fps.

Fig. 5 shows the relative occurrence of the multihypothesis mode over the quantization parameter for the CIF sequence *Mobile & Calendar*. Results for a long-term memory buffer size of $M = 1, 2, 5,$ and 10 are plotted. For high bit-rates (small quantization parameters), the multihypothesis mode exceeds a relative occurrence of 50% among all H.26L INTER-modes. In addition, this occurrence increases for larger long-term memory buffer sizes and correlates to improved compression efficiency.

4. EXPERIMENTAL RESULTS

Our coder is based on the H.26L test model TML-6 [6]. For our experiments, the CIF sequences *Mobile & Calendar* and *Tempete* are coded at 30 fps. We investigate the rate-distortion performance of multihypothesis pictures with respect to H.26L P pictures for various long-term memory buffer sizes.

Figs. 6 and 7 show the bit-rate values at 35 dB PSNR of the luminance signal over the number of reference frames M for the sequences *Mobile & Calendar* and *Tempete*, respectively. We compute PSNR vs. bit-rate curves by varying the quantization parameter and interpolate intermediate points by a cubic spline. The performance of H.26L P pictures (TML) and the corresponding multihypothesis extension with linearly combined macrohypotheses (TML+MHP) is shown.

The multihypothesis codec with $M = 1$ reference frames has to choose both macrohypotheses from the previous frame. For $M > 1$, we allow more than one reference frame for each macrohypothesis. The reference frames for both hypotheses are selected by the rate-constrained multihypothesis motion estimation algorithm described in section 3. The picture reference parameter allows also the special case that both macrohypotheses are chosen from the same reference frame. The rate constraint is responsible for the trade-off between prediction quality and bit-rate. Turning on the multihypothesis extension for the codec with $M = 10$ reference fames, the bit-rate is reduced from 2019 to 1750 kbit/s when coding the sequence *Mobile & Calendar*. This corresponds to 13% bit-rate savings. The gain by the multihypothesis codec with just one reference frame is limited to 6%. The gain by the multihypothesis codec over the TML improves for a increasing number of reference frames. This observation is independent of the implemented multihypothesis prediction scheme [5].

Figs. 8 and 9 depict the average luminance PSNR from reconstructed frames over the overall bit-rate produced by H.26L P pictures (TML) and the multihypothesis codec (TML+MHP) for the sequences *Mobile & Calendar* and *Tempete*. The number of reference frames is chosen to be $M = 1$ and $M = 5$. It can be observed that the compression efficiency improves for increasing bit-rate.

5. CONCLUSIONS

Multihypothesis pictures extend the H.26L P pictures and achieve additional compression efficiency. The H.26L syntax for the bidirectional mode in B pictures allows a linear combination of two macrohypotheses. With multihypothesis pictures, the gain of linearly combined macrohypotheses is not only restricted to inserted B pictures; all coded pictures can benefit from linearly combined motion-compensated prediction signals. MH pictures utilize temporally previous pictures and cause therefore no extra coding delay. The utilized iterative algorithm for multihypothesis motion estimation has low complexity and is based on standard techniques for motion estimation. It takes about 2 iterations to determine an efficient macrohypothesis pair. We observed bit-rate savings up to 13% for MH pictures over H.26L P pictures with 5 reference frames. The gain by multihypothesis prediction over single hypothesis prediction improves for an increasing number of reference frames. That is, multiple reference frames allow efficient macrohypothesis pairs.

6. REFERENCES

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

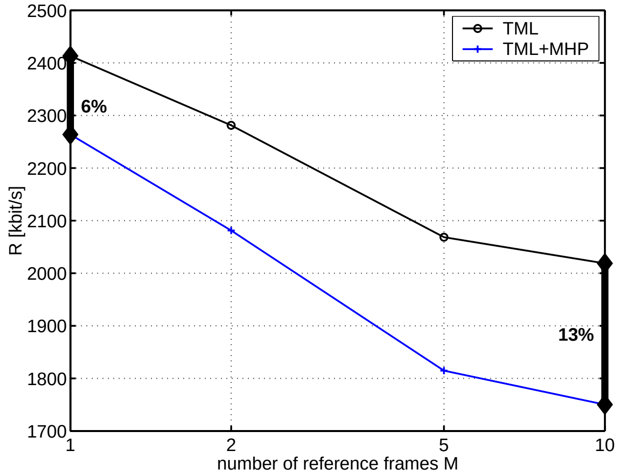

Fig. 6. Avg. bit-rate at 35 dB PSNR vs. number of reference frames for the CIF sequence *Mobile & Calendar* with 30 fps.

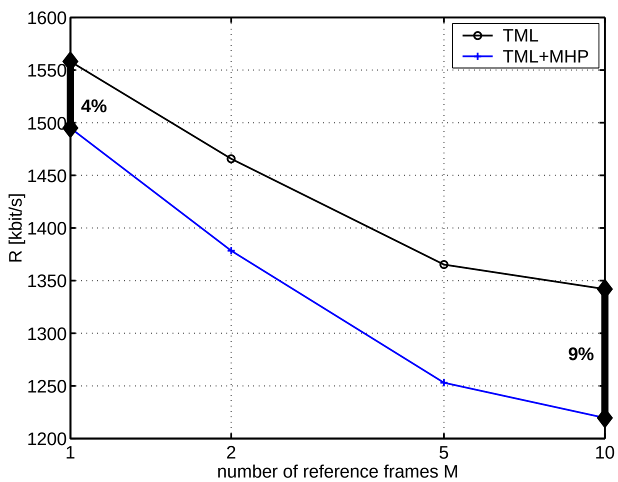

Fig. 7. Avg. bit-rate at 35 dB PSNR vs. number of reference frames for the CIF sequence *Tempete* with 30 fps.

[2] M. Flierl, T. Wiegand, and B. Girod, "Rate-Constrained Multi-Hypothesis Motion-Compensated Prediction for Video Coding," in *Proceedings of the IEEE International Conference on Image Processing*, Vancouver, Canada, Sept. 2000, vol. III, pp. 150–153.

[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] M. Flierl and B. Girod, "Multihypothesis Motion Estimation for Video Coding," in *Proceedings of the Data Compression Conference*, Snowbird, Utah, Mar. 2001, pp. 341–350.

[5] M. Flierl and B. Girod, "Multihypothesis Motion-Compensated Prediction with Forward-Adaptive Hypothesis Switching," in *Proceedings of the Picture Coding Symposium*, Seoul, Korea, Apr. 2001, pp. 195–198.

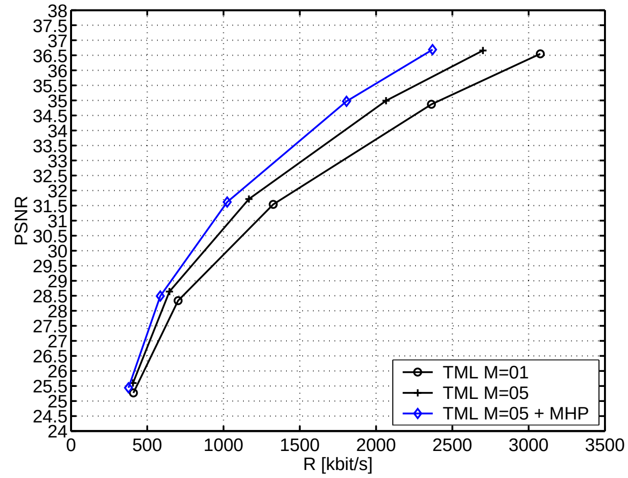

Fig. 8. PSNR of the luminance signal vs. overall bit-rate for the CIF sequence *Mobile & Calendar* with 30 fps.

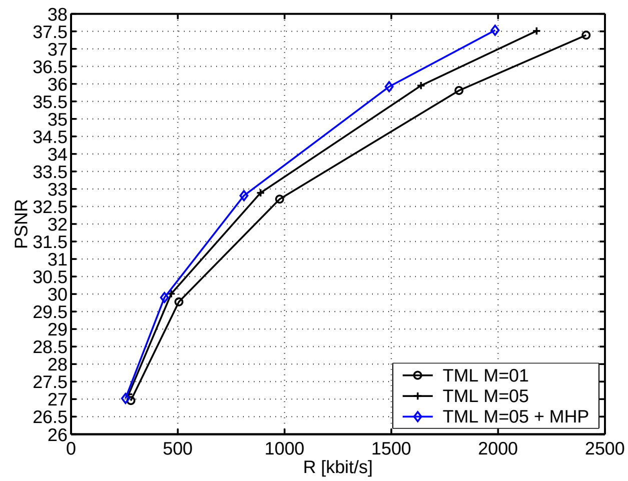

Fig. 9. PSNR of the luminance signal vs. overall bit-rate for the CIF sequence *Tempete* with 30 fps.

[6] ITU-T Video Coding Experts Group, *H.26L Test Model Long Term Number 6*, Mar. 2001, Download via anonymous ftp to: standard.pictel.com/video-site/0101_Eib/ VCEG-L45d0.doc.

[7] M. Flierl and B. Girod, *Further Investigation of Multihypothesis Motion Pictures*, ITU-T Video Coding Experts Group, Apr. 2001, Download via anonymous ftp to: standard.pictel.com/video-site/0401_Aus/ VCEG-M40.doc.

[8] 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.

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