Generalized B Pictures and the Draft H.264/AVC Video Compression Standard

Markus Flierl, *Student Member, IEEE*, and Bernd Girod, *Fellow, IEEE*

Invited Paper

Abstract— **This paper reviews recent advances in using B pictures in the context of the draft H.264/AVC video compression standard. We focus on reference picture selection and linearly combined motion-compensated prediction signals. We show that bi-directional prediction exploits partially the efficiency of combined prediction signals whereas multihypothesis prediction allows a more general form of B pictures. The general concept of linearly combined prediction signals chosen from an arbitrary set of reference pictures improves the H.264/AVC test model TML-9 which is used in the following.**

We outline H.264/AVC macroblock prediction modes for B pictures, classify them into four groups and compare their efficiency in terms of rate-distortion performance. When investigating multihypothesis prediction, we show that bi-directional prediction is a special case of this concept. Multihypothesis prediction allows also two combined forward prediction signals. Experimental results show that this case is also advantageous in terms of compression efficiency.

The draft H.264/AVC video compression standard offers improved entropy coding by context-based adaptive binary arithmetic coding. Simulations show that the gains by multihypothesis prediction and arithmetic coding are additive.

B pictures establish an enhancement layer and are predicted from reference pictures that are provided by the base layer. The quality of the base layer influences the rate-distortion trade-off for B pictures. We demonstrate how the quality of the B pictures should be reduced to improve the overall rate-distortion performance of the scalable representation.

Keywords— **Video Coding, Motion-Compensated Prediction, Multihypothesis Motion-Compensated Prediction, B Pictures, Multiframe Prediction, Temporal Scalability**

I. Introduction

B PICTURES are pictures in a motion video sequence that are encoded using both past and future pictures as references. The prediction is obtained by a linear combination of forward and backward prediction signals usually obtained with motion compensation. However, such a superposition is not necessarily limited to forward and backward prediction signals [1], [2]. For example, a linear combination of two forward prediction signals can also be efficient in terms of compression efficiency. The prediction method which linearly combines motion-compensated signals regardless of the reference picture selection will be referred to as multihypothesis motion-compensated prediction [3]. The concept of reference picture selection [4], also called multiple reference picture prediction, is utilized to allow prediction from both temporal directions. In this particular case, a bi-directional picture reference parameter addresses both past and future reference pictures [5]. This generalization in terms of picture reference selection and linearly combined prediction signals is reflected in the term *generalized B pictures* and is realized in the emerging H.264/AVC video compression standard [6]. It is desirable that an arbitrary pair of reference pictures can be signaled to the decoder [7], [8]. This includes the classical combination of forward and backward prediction signals but also allows forward/forward as well as backward/backward pairs. When combining the two most recent pictures, a functionality similar to the dual-prime mode in MPEG-2 [9], [10] is achieved, where top and bottom fields are averaged to form the final prediction.

The efficiency of forward and backward prediction has already been raised by Musmann et al. in 1985 [11]. In the same year, Ericsson [12] published investigations on adaptive predictors for hybrid coding that use up to four previous fields. A rate-distortion efficient technique for block-based reference picture selection was introduced by Wiegand et al. [4]. The now known concept of B pictures was proposed to MPEG by Puri et al. [13]. The motivation was to interpolate any skipped frame taking into account the movement between the two 'end' frames. The technique, called conditional motion-compensated interpolation, coupled the motion-compensated interpolation strategy with transmission of the significant interpolation errors.

A theoretical analysis of multihypothesis motion-compensated prediction in [3] discusses performance bounds for hybrid video coding: In the noiseless case, increasing the accuracy of motion compensation from, e.g., half-pel to quarter-pel reduces the bit-rate of the residual encoder by at most 1 bit/sample. In the case of uncorrelated displacement errors, doubling the number of linearly combined prediction signals gains at most 0.5 bits/sample. The overall performance of motion-compensated prediction is limited by the residual noise which is also lowered by linearly combined prediction signals. [14] investigates optimal multihypothesis motion estimation. It is demonstrated that joint estimation of several motion-compensated signals implies maximally negatively correlated displacement errors. In the noiseless case, increasing the accuracy of multihypothesis motion-compensated prediction from, e.g., half-pel to quarter-pel reduces the bit-rate of the residual encoder

M. Flierl visited the Information Systems Laboratory at Stanford University, Stanford, CA, and is on leave from the Telecommunications Laboratory, University of Erlangen-Nuremberg, Erlangen, Germany. e-mail: mflierl@ieee.org

B. Girod is with the Information Systems Laboratory, Stanford University, Stanford, CA. e-mail: bgirod@stanford.edu

by at most 2 bits/sample. This improvement is already observed for two hypotheses and also applies to predictors of higher order. With respect to multiple reference picture prediction, doubling the number of reference pictures for motion-compensated prediction reduces the bit-rate of the residual encoder by at most 0.5 bits/sample. Whereas doubling the number of reference pictures for multihypothesis motion-compensated prediction reduces the bit-rate of the residual encoder by at most 1 bit/sample [15].

B pictures in H.264/AVC have been improved in several ways compared to B pictures in MPEG-2 [9] and H.263 [16]. The block size for motion compensation can range from 16×16 to 4×4 pixels and the direct mode with weighted blending allows not only a scaling of the motion vectors but also a weighting of the prediction signal. The current draft standard H.264/AVC provides also improved H.263 Annex U functionality.

H.263 Annex U, "Enhanced Reference Picture Selection", already allows multiple reference pictures for forward prediction and two-picture backward prediction in B pictures. When choosing between the most recent and the subsequent reference picture, the multiple reference picture selection capability is very limited. Utilizing multiple prior and subsequent reference pictures improves the compression efficiency of H.263 B pictures.

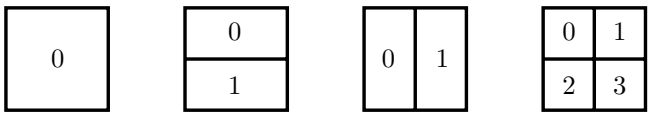
The H.264/AVC test model software TML-9 [17] uses only inter pictures as reference pictures to predict the B pictures. B pictures can be utilized to establish an enhancement layer in a layered representation and allow temporal scalability [18]. That is, decoding of a sequence at more than one frame rate is achievable. In addition to this functionality, B pictures generally improve the overall compression efficiency when compared to that of inter pictures only [19]. On the other hand, they increase the time delay due to multiple future reference pictures. But this disadvantage is not critical in applications like Internet streaming and multimedia storage for entertainment purpose. Beyond the test model software TML-9, and different from past standards, the multiple reference picture framework RPSL in H.264/AVC also allows previously decoded B pictures to be used as reference for B picture coding [20].

In the following, we present selected features of the B pictures in the draft H.264/AVC video compression standard with possible extensions and investigate their performance in terms of compression efficiency. Section II introduces B picture prediction modes. After an overview, direct and multihypothesis mode are discussed in more detail and a rate-distortion performance comparison of three mode classes is provided. Section III elaborates multihypothesis prediction. The difference between bi-directional and multihypothesis mode is outlined and quantified in experimental results. In addition, the efficiency of two combined forward prediction signals is also investigated. Finally, both H.264/AVC entropy coding schemes are investigated with respect to the multihypothesis mode. Encoder issues are detailed in Section IV, which covers rate-constrained mode decision, motion estimation, and multihypothesis motion estimation. In addition, the improvement of the overall rate-distortion performance with B pictures is discussed.

II. Prediction Modes for B Pictures

A. Overview

The macroblock modes for B pictures allow intra and inter coding. The **intra-mode macroblocks** specified for inter pictures are also available for B pictures. The **inter-mode macroblocks** are especially tailored to B picture use. As for inter pictures, they utilize four generic partitions as depicted in Fig. 1 to generate the motion-compensated macroblock prediction signal. The current draft [6] combines this generic partition with a hierarchical framework[1] which permits block sizes up to 4×4. In addition, the use of the reference picture set available for predicting the current B picture is suited to its temporally non-causal character.



16×16 8×8 | 16×8 8×4 | 8×16 4×8 | 8×8 4×4

Fig. 1. Macroblock and sub-macroblock partitions for the motion-compensated prediction signal.

In contrast to the previously mentioned inter-mode macroblocks which signal motion vector data according to its block size as side-information, the **direct-mode macroblock** does not require such side-information but derives reference frame, block size, and motion vector data from the subsequent inter picture. This mode superimposes two prediction signals. One prediction signal is derived from the subsequent inter picture, the other from a previous picture.

A linear combination of two motion-compensated prediction signals with explicit side-information is accomplished by the **multihypothesis-mode macroblock**. Existing standards with B pictures utilize the bi-directional mode, which only allows the combination of a previous and subsequent prediction signal. The multihypothesis mode generalizes this concept and supports not only the already mentioned forward/backward prediction pair, but also forward/forward and backward/backward pairs.

B. Direct Mode

The direct mode uses bi-directional prediction and allows residual coding of the prediction error. The forward and backward motion vectors (MV_0, MV_1) of this mode are derived from the motion vectors MV_C used in the co-located macroblock of the subsequent picture RL_1. Note that the direct-mode macroblock uses the same partition as the co-located macroblock. The prediction signal is calculated by a linear combination of two blocks that are determined by the forward and backward motion vectors pointing to two reference pictures (RL_0, RL_1). When using multiple reference picture prediction, the forward reference picture for

[1]This is a generalization of the seven block size types that have been used for the test model TML-9 [17].

the direct mode RL_1 is chosen to be the subsequent inter picture with the co-located macroblock; see Fig. 2.

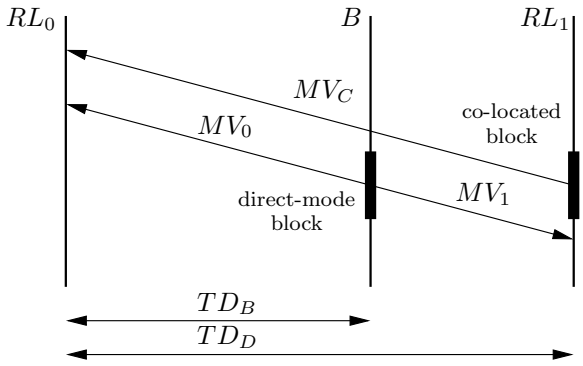


Fig. 2. A direct-mode block has two derived motion vectors MV_0 and MV_1 pointing to two reference pictures RL_0 and RL_1.

The forward and backward motion vectors for direct-mode blocks are calculated as

$$MV_0 = \frac{TD_B}{TD_D} MV_C \quad (1)$$

$$MV_1 = \frac{TD_B - TD_D}{TD_D} MV_C, \quad (2)$$

where MV_0 is the forward motion vector, MV_1 is the backward motion vector, and MV_C represents the motion vector of the co-located block in the subsequent inter picture. For classic B pictures, TD_D is the temporal distance between the previous and the next inter picture, and TD_B is the distance between the current B picture and the previous inter picture. When multiple reference picture prediction is in use, the current draft [6] uses modified definitions for the temporal distances. In that case, the actual reference picture RL_0 (which is also reference picture for the co-located macroblock of the following picture) is used for the calculation of the temporal distances TD_D and TD_B. And when both the current macroblock and its co-located are in frame mode, TR_B is the temporal distance between the current B frame and the reference frame RL_0, and TR_D is the temporal distance between the subsequent reference frame $RL1$ and RL_0.

The direct mode in H.264/AVC is improved by weighted blending of the prediction signal [21]. Video content like music videos and movie trailers make frequent use of fading transitions from scene to scene. It is very popular in movie trailers to fade each scene to black, and then from black to the next scene. Without weighted blending of the prediction signal, both normal fades and "fades to-black" are hard to encode well without visible compression artifacts. For example, when encoding with a PBBB pattern, the B pictures in position 1 and 3 suffer from quality degradation relative to the B pictures in position 2 and the surrounding inter and intra pictures. The weighted blending technique considers how the direct mode motion vectors are derived from scaling the motion vector for the subsequent inter picture, based on the distance between the B picture and the surrounding pictures, and also weighs the calculation of the prediction block based on this distance, instead of the averaging with equal weights that has been used in all existing standards with B pictures. The weighted blending technique calculates the prediction block c for direct mode coded macroblocks according to

$$c = \frac{c_p(TD_D - TD_B) + c_s TD_B}{TD_D}, \quad (3)$$

where c_p is the prediction block from a previous reference picture, and c_s is the prediction block from the subsequent reference picture. Sequences without any fades will not suffer from loss of compression efficiency relative to the conventional way of calculating the prediction for the direct mode.

C. Multihypothesis Mode

The multihypothesis mode superimposes two macroblock prediction signals with their individual sets of motion vectors. We refer to each block prediction signal as a "hypotheses". To calculate prediction blocks, the motion vectors of the two hypotheses are used to obtain appropriate blocks from reference pictures. The obtained blocks are just averaged. For TML-9 [17], each macroblock hypothesis is specified by one of the seven block size types and one picture reference parameter. But the current draft of H.264/AVC uses the generic partition according to Fig. 1 which includes the seven partitions as used in TML-9. In addition, the current draft assigns the reference picture parameter at the block level. It is very likely that the hypotheses are chosen from different reference pictures but they can also originate from the same picture. Increasing the number of available reference pictures actually improves the performance of multihypothesis motion-compensated prediction [15]. More details are given in Section III.

D. Rate-Distortion Performance of Individual Modes

The macroblock modes for B pictures can be classified into four groups:

1. No extra side-information is transmitted for this particular macroblock. This corresponds to the **direct** mode.
2. Side-information for one macroblock prediction signal is transmitted. The **inter** modes with block structures according to Fig. 1 and bi-directional picture reference parameters belong to this group.
3. Side-information for two macroblock prediction signals is transmitted to allow **multihypothesis** prediction.
4. The last group includes all **intra** modes and excludes any kind of inter-frame prediction.

In the following, the first three groups which utilize inter-frame prediction are investigated with respect to their rate-distortion performance. The fourth group with the intra modes is used for encoding but is not discussed further.

The rate-distortion performance of the groups **direct, inter,** and **MH** are depicted in Fig. 3. The PSNR of the B

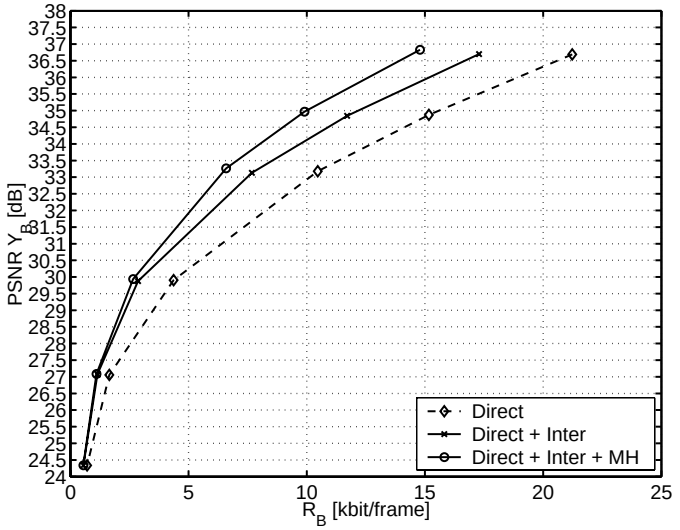


Fig. 3. PSNR of the B picture luminance signal vs. B picture bit-rate for the QCIF sequence *Mobile & Calendar* with 30 fps. Two B pictures are inserted after each inter picture. 5 past and 3 subsequent reference pictures are used. The compression efficiency of the B picture coding modes *direct, inter,* and *MH* are compared.

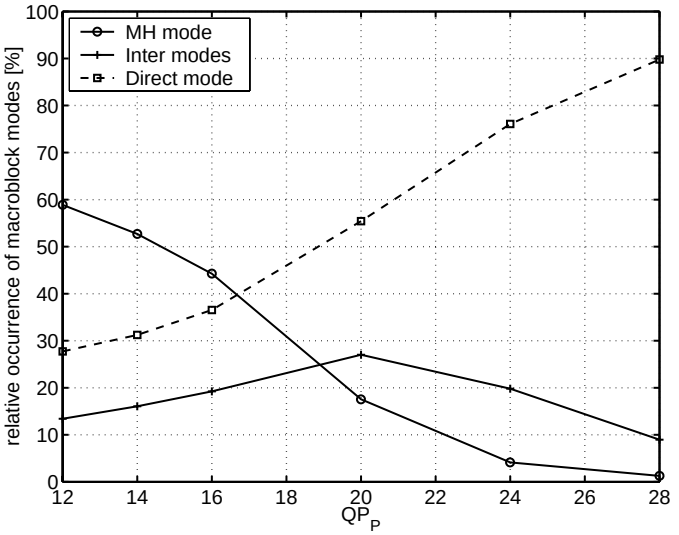


Fig. 4. Relative occurrence of the macroblock modes in B pictures vs. quantization parameter for the QCIF sequence *Mobile & Calendar* with 30 fps. Two B pictures are inserted after each inter picture. 5 past and 3 subsequent reference pictures are used. The relative frequency of the B picture macroblock modes *direct, inter,* and *MH* are compared.

picture luminance signal is plotted over the B picture bit-rate for the QCIF sequence *Mobile & Calendar*. With the direct mode for the B pictures, the rate-distortion performance at high bit-rates is dominated by the efficiency of the residual encoding. The inter modes improve the compression efficiency approximately by 1 dB in PSNR at moderate and high bit-rates. At very low bit-rates, the rate-penalty in effect disables the modes in the inter group due to extra side-information. Similar behavior can be observed for the multihypothesis (MH) mode. Transmitting two prediction signals increases the side-information additionally. Consequently, the multihypothesis mode improves compression efficiency approximately by 1 dB in PSNR at high bit-rates.

Corresponding to the rate-distortion performance of the three groups, Fig. 4 depicts the relative occurrence of the macroblock modes in B pictures vs. quantization parameter QP_P for the QCIF sequence *Mobile & Calendar*. At $QP_P = 28$ (low bit-rate), the direct mode is dominant with more than 90 %, whereas the multihypothesis and inter modes are suppressed by the rate constraint. At $QP_P = 12$, the relative occurrence of the direct mode decreases to 30 %, whereas the relative frequency of the multihypothesis mode increases to 60 %. About 10 % of the macroblocks utilize an inter mode.

The influence of the B picture coding modes *direct, inter,* and *MH* on the overall compression efficiency is depicted in Fig. 5 for the QCIF sequence *Mobile & Calendar*. The base layer (the sequence of inter pictures) is identical in all three cases and only the B picture coding modes are selected from the specified classes. For this sequence, the inter modes in the B pictures improve the overall efficiency approximately by 0.5 dB. The multihypothesis mode adds an additional 0.5 dB for higher bit-rates.

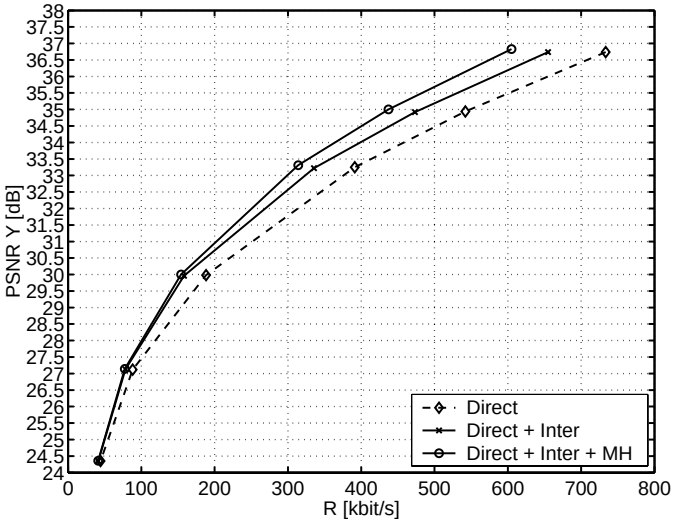


Fig. 5. PSNR of the luminance signal vs. overall bit-rate for the QCIF sequence *Mobile & Calendar* with 30 fps. Two B pictures are inserted after each inter picture. 5 past and 3 subsequent reference pictures are used. The compression efficiency of the B picture coding modes *direct, inter,* and *MH* are compared.

III. Multihypothesis Prediction

A. Bi-Directional vs. Multihypothesis Mode

In the following, we will outline the difference between the bi-directional macroblock mode, which is specified in the H.264/AVC test model TML-9 [17], and the multihypothesis mode proposed in [8] and discussed in the previous section. A bi-directional prediction type only allows a linear combination of a forward / backward prediction pair; see Fig. 6. The test model TML-9 utilizes multiple reference pictures for forward prediction but allows only backward prediction from the most subsequent reference picture. For

bi-directional prediction, independently estimated forward and backward prediction signals are practical but the efficiency can be improved by joint estimation. For multihypothesis prediction in general, a joint estimation of two hypotheses is necessary [14]. An independent estimate might even deteriorate the performance. The test model software TML-9 does not allow a joint estimation of forward and backward prediction signals.

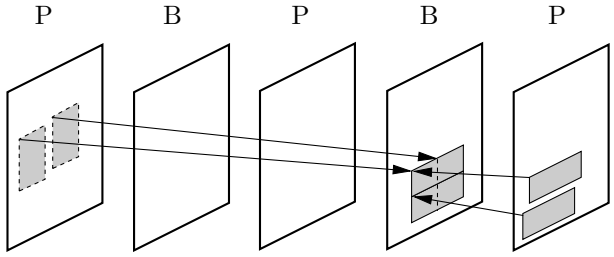


Fig. 6. A bi-directional prediction mode allows a linear combination of one past and one subsequent macroblock prediction signal. The inter pictures are denoted by P.

The multihypothesis mode removes the restriction of the bi-directional mode to allow only linear combinations of forward and backward pairs. The additional combinations (forward, forward) and (backward, backward) are obtained by extending an unidirectional picture reference syntax element to a bi-directional picture reference syntax element; see Fig. 7.

With this bi-directional picture reference element, a generic prediction signal, which we call hypothesis, can be formed with the syntax fields for reference frame, block size, and motion vector data.

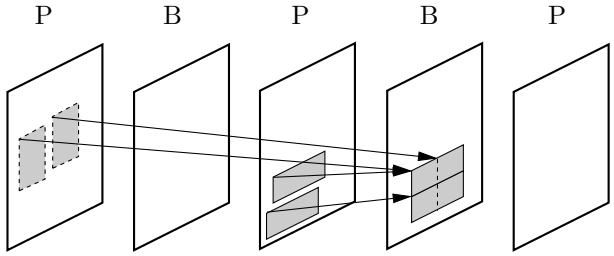


Fig. 7. The multihypothesis mode also allows a linear combination of two past macroblock prediction signals. The inter pictures are denoted by P.

The multihypothesis mode includes the bi-directional prediction mode when the first hypothesis originates from a past reference picture and the second from a future reference picture. The bi-directional mode limits the set of possible reference picture pairs. Not surprisingly, a larger set of reference picture pairs improves the coding efficiency of B pictures.

The following results are based on the H.264/AVC test model TML-9 [17]. For our experiments, the CIF sequences *Mobile & Calendar* and *Flowergarden* are coded at 30 fps. We investigate the rate-distortion performance of the multihypothesis mode in comparison with the bi-directional mode when two B pictures are inserted.

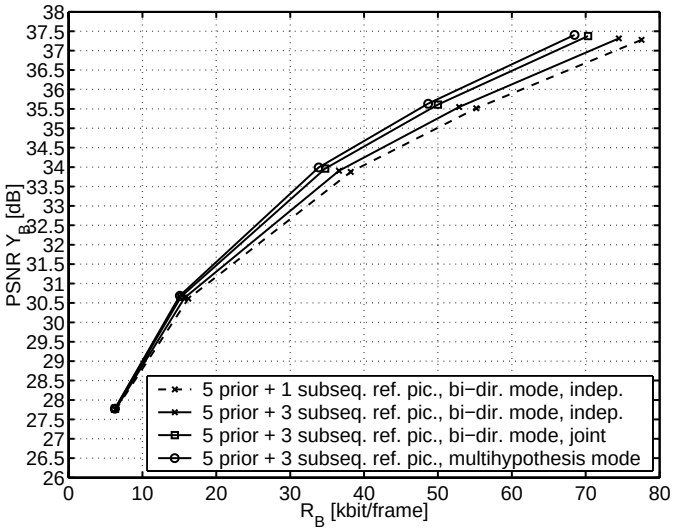


Fig. 8. PSNR of the B picture luminance signal vs. B picture bit-rate for the CIF sequence *Mobile & Calendar* with 30 fps. Two B pictures are inserted after each inter picture. $QP_B = QP_P$. The multihypothesis mode is compared to the bi-directional mode.

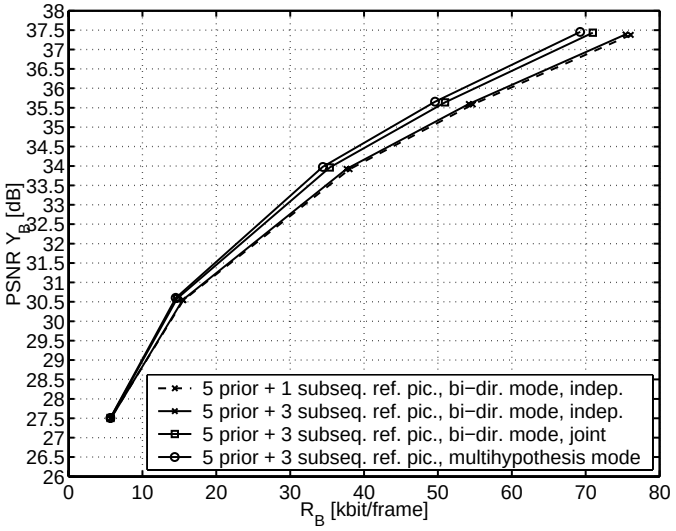


Fig. 9. PSNR of the B picture luminance signal vs. B picture bit-rate for the CIF sequence *Flowergarden* with 30 fps. Two B pictures are inserted after each inter picture. $QP_B = QP_P$. The multihypothesis mode is compared to the bi-directional mode.

Figs. 8 and 9 depict the average luminance PSNR from reconstructed B pictures over the overall bit-rate produced by B pictures with bi-directional prediction mode and the multihypothesis mode for the sequences *Mobile & Calendar* and *Flowergarden*. The number of reference pictures is chosen to be 1 and 3 future reference pictures with a constant number of 5 past pictures. It can be observed that increasing the total number of reference pictures from $5 + 1$ to $5 + 3$ slightly improves compression efficiency. Moreover, the multihypothesis mode outperforms the bi-directional mode and its compression efficiency improves for increasing bit-rate. In the case of the bi-directional mode, jointly estimated forward and backward prediction signals outperform independently estimated signal pairs.

Please note that linearly combined prediction signals not only take advantage of suppressed noise components but also provide occlusion benefits, in particular for bi-directional prediction. Let us assume objects that appear in the current frame as well as in future frames but are occluded in past frames. It is likely that more reference pictures from the past will not be able to improve the prediction efficiency as the objects are occluded. But several future reference pictures are likely to predict this object more efficiently.

B. Two Combined Forward Prediction Signals

Generalized B pictures combine both the superposition of prediction signals and the reference picture selection from past and future pictures. In the following, we investigate generalized B pictures with forward-only prediction and utilize them like inter pictures for comparison purposes [7]. That is, only a unidirectional reference picture parameter which addresses past pictures is permitted. As there is no future reference picture, the direct mode is replaced by the skip mode as specified for inter pictures. The *generalized B pictures with forward-only prediction* cause no extra coding delay as they utilize only past pictures for prediction and are also used for reference to predict future pictures.

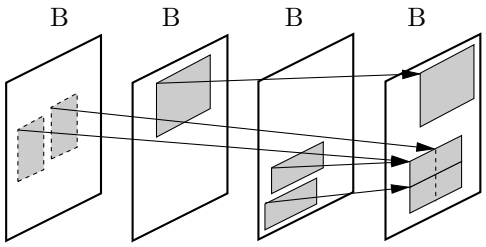


Fig. 10. Generalized B pictures with forward-only prediction utilize multiple reference picture prediction and multihypothesis motion-compensated prediction. The multihypothesis mode uses two hypotheses chosen from past reference pictures.

Fig. 10 shows generalized B pictures with forward-only prediction. They allow multiple reference picture prediction and linearly combined motion-compensated prediction signals with individual block size types. Both hypotheses are just averaged to form the current macroblock. The test model TML-9 [17] allows seven different block sizes which will be the seven hypotheses types in the multihypothesis mode. For inter modes, TML-9 allows only one picture reference parameter per macroblock and assumes that all sub-blocks can be found on that specified reference picture. But the current draft H.264/AVC is similar to the H.263 standard, where multiple reference picture prediction utilizes picture reference parameters for 16×16 and 8×8 block shapes.

We investigate the rate-distortion performance of generalized B pictures with forward-only prediction and compare them to H.264/AVC inter pictures for various numbers of reference pictures. Fig. 11 shows the bit-rate values at 35 dB PSNR of the luminance signal over the number of reference pictures M for the CIF sequences *Mobile & Calendar*, coded at 30 fps. We compute PSNR vs. bit-rate curves by varying the quantization parameter and interpolate intermediate points by a cubic spline. The performance of H.264/AVC inter pictures (IPPP...) and the generalized B pictures with forward-only prediction (IBBB...) is shown.

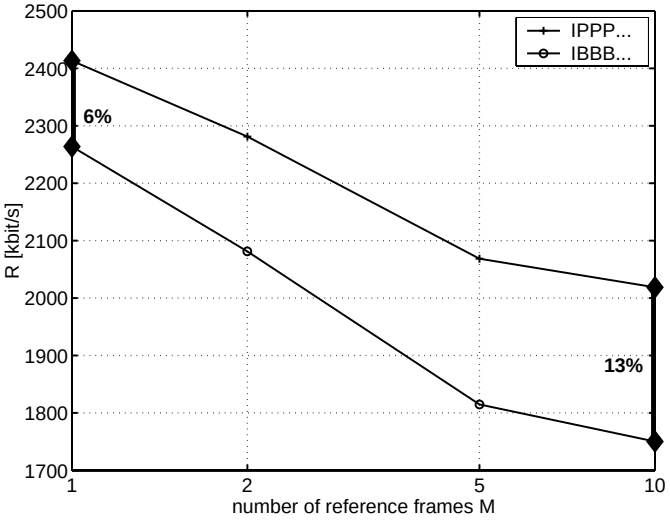


Fig. 11. Average bit-rate at 35 dB PSNR vs. number of reference pictures for the CIF sequence *Mobile & Calendar* with 30 fps. Generalized B pictures with forward-only prediction are compared to inter pictures.

The generalized B pictures with forward-only prediction and $M = 1$ reference picture has to choose both hypotheses from the previous picture. For $M > 1$, we allow more than one reference picture for each hypothesis. The reference pictures for both hypotheses are selected by the rate-constrained multihypothesis motion estimation algorithm described in Section IV-C. The picture reference parameter allows also the special case that both hypotheses are chosen from the same reference picture. The rate constraint is responsible for the trade-off between prediction quality and bit-rate. Using the generalized B pictures with forward-only prediction and $M = 10$ reference pictures reduces the bit-rate from 2019 to 1750 kbit/s when coding the sequence *Mobile & Calendar*. This corresponds to 13% bit-rate savings. The gain by the generalized B pictures with forward-only prediction and just one reference picture is limited to 6%. The gain by the generalized B pictures over the inter pictures improves for a increasing number of reference pictures [22]. This observation is independent of the implemented multihypothesis prediction scheme [15].

Fig. 12 depicts the average luminance PSNR from reconstructed pictures over the overall bit-rate produced by TML-9 inter pictures (IPPP...) and the generalized B pictures with forward prediction only (IBBB...) for the sequences *Mobile & Calendar*. The number of reference pictures is chosen to be $M = 1$ and $M = 5$. It can be observed that the gain by generalized B pictures improves for increasing bit-rate.

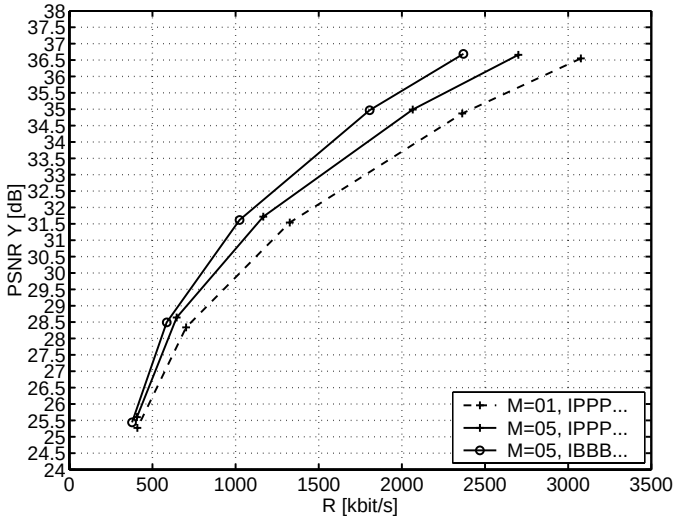


Fig. 12. PSNR of the luminance signal vs. overall bit-rate for the CIF sequence *Mobile & Calendar* with 30 fps. Generalized B pictures with forward-only prediction are compared to inter pictures.

C. Entropy Coding

Entropy coding for TML-9 B pictures can be carried out in one of two different ways: universal variable length coding (UVLC) or context-based adaptive binary arithmetic coding (CABAC) [23], [24], [25]. The UVLC scheme uses only one variable length code to map all syntax elements to binary representations whereas CABAC utilizes context modeling and adaptive arithmetic codes to exploit conditional probabilities and non-stationary symbol statistics [24]. The simplicity of the UVLC scheme is striking as it demonstrates good compression efficiency at very low computational costs. CABAC with higher computational complexity provides additional bit-rate savings mainly for low and high bit-rates. A detailed discussion of CABAC is presented in [26].

The syntax elements used by the multihypothesis mode can be coded with both the UVLC and the CABAC scheme. When using CABAC for the multihypothesis mode, the context model for motion vector data is adapted to multihypothesis motion. The utilized context model $ctx\_mvd(C,k)$ for the difference motion vector component $mvd_k(C)$ and the current block C is

$$ctx\_mvd(C,k) = \begin{cases} 0 & \text{for} \quad e_k(C) < 3, \\ 1 & \text{for} \quad e_k(C) > 15, \\ 2 & \text{else}, \end{cases} \tag{4}$$

where $e_k(C)$ captures the motion activity of the context. The difference motion vector of the current block $mvd(C)$ is the difference between the estimated motion vector of the current block and a predicted motion vector obtained from spatial neighbors. For the first hypothesis, $e_k(C)$ is the sum of the magnitude of difference motion vector components from neighboring blocks to the left and to the top

$$e_k(C) = |mvd_k(\text{left})| + |mvd_k(\text{top})|. \tag{5}$$

For the second hypothesis, $e_k(C)$ is the absolute value of the difference motion vector component mvd_k of the first hypothesis

$$e_k(C) = |mvd_k(\text{first hypothesis})|. \tag{6}$$

The context models for the remaining syntax elements are not altered. Experimental results show that generalizing the bi-directional mode to the multihypothesis mode improves B picture compression efficiency not only for the UVLC scheme. It will be shown that the gains by the multihypothesis mode and the CABAC scheme are additive.

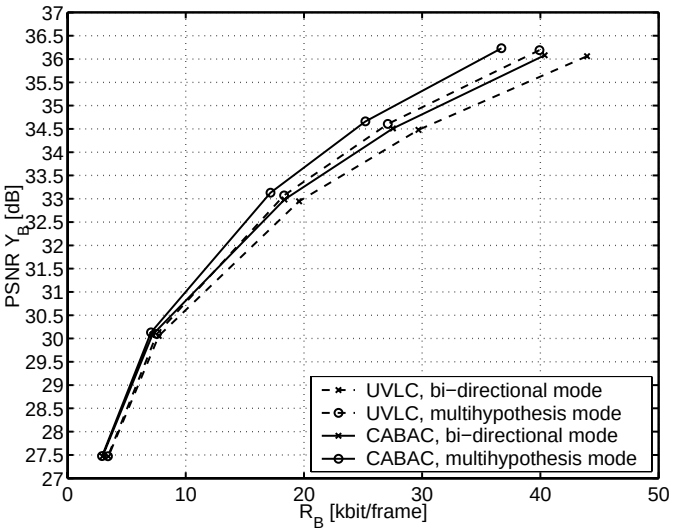


Fig. 13. PSNR of the B picture luminance signal vs. B picture bit-rate for the CIF sequence *Mobile & Calendar* with 30 fps. Two B pictures are inserted after each inter picture. 5 past and 3 future inter pictures are used for predicting each B picture. $QP_B = QP_P + 2$ and $\lambda_B = 4f(QP_P)$. The multihypothesis mode and the bi-directional mode with independent estimation are compared for both entropy coding schemes.

Figs. 13 and 14 depict the B picture compression efficiency for the CIF sequences *Mobile & Calendar* and *Flowergarden*, respectively. For motion-compensated prediction, 5 past and 3 future inter pictures are used in all cases. The multihypothesis mode and the bi-directional mode with independent estimation of prediction signals are compared for both entropy coding schemes. The PSNR gains by the multihypothesis mode and the CABAC scheme are somewhat comparable for the investigated sequences at high bit-rates. When enabling the multihypothesis mode with CABAC, additive gains can be observed. The multihypothesis mode improves the efficiency of motion-compensated prediction and CABAC optimizes the entropy coding of the utilized syntax elements.

IV. Encoder Issues

A. Rate-Constrained Mode Decision

The test model TML-9 distinguishes between a low- and high-complexity encoder. For a low-complexity encoder, computationally inexpensive rules for mode decision are recommended [27]. For a high-complexity encoder, the macroblock mode decision is ruled by minimizing the La-

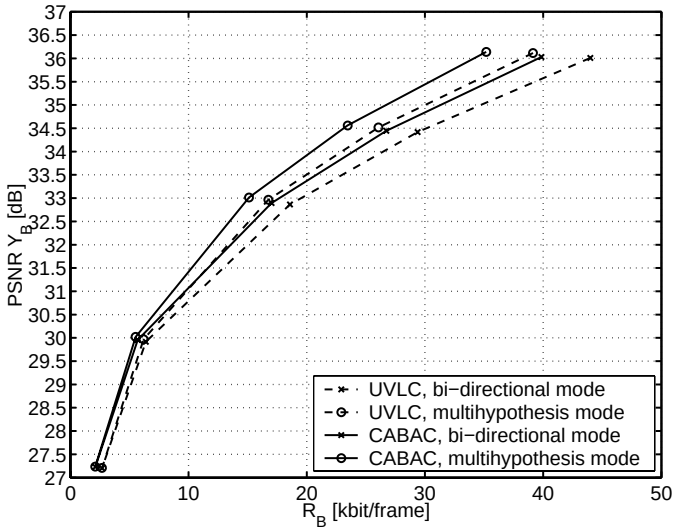


Fig. 14. PSNR of the B picture luminance signal vs. B picture bit-rate for the CIF sequence *Flowergarden* with 30 fps. Two B pictures are inserted after each inter picture. 5 past and 3 future inter pictures are used for predicting each B picture. $QP_B = QP_P + 2$ and $\lambda_B = 4f(QP_P)$. The multihypothesis mode and the bi-directional mode with independent estimation are compared for both entropy coding schemes.

grangian function

$$J_1(\text{Mode} \mid QP, \lambda) = SSD(\text{Mode} \mid QP) + \lambda R(\text{Mode} \mid QP), \tag{7}$$

where QP is the macroblock quantizer parameter, and λ the Lagrange multiplier for mode decision. *Mode* indicates the selection from the set of potential coding modes. SSD is the sum of the squared differences between the original block and its reconstruction. It also takes into account the distortion in the chrominance components. R is the number of bits associated with choosing *Mode* and QP, including the bits for macroblock header, motion information, and all integer transform blocks. The Lagrangian multiplier for λ is related to the macroblock quantizer parameter QP by

$$\lambda := f(QP) = 5\frac{QP+5}{34-QP}\exp\left(\frac{QP}{10}\right). \tag{8}$$

Detailed discussions of this relationship can be found in [28] and [29]. Experimental results in Section IV-D verify that this relation should be adapted for B pictures as specified in the test model TML-9,

$$\lambda_B = 4f(QP_B), \tag{9}$$

such that the overall rate-distortion efficiency for the sequence is improved.

Mode decision selects the best mode among all B picture macroblock modes and captures both prediction and prediction error encoding. Note that prediction error encoding is dependent on the performance of the predictor. In general, this dependency requires joint encoding. But it is practical to determine the prediction parameters with rate-constrained motion estimation independent of the prediction error encoding. A detailed discussion of the rate-constrained coder control is presented in [30].

B. Rate-Constrained Motion Estimation

Motion estimation is also performed in a rate-constrained framework. The encoder minimizes the Lagrangian cost function

$$J_2(m, r \mid \lambda_{\text{SAD}}, p) = SAD(m, r) + \lambda_{\text{SAD}} R(m - p, r), \tag{10}$$

with the motion vector m, the predicted motion vector p, the reference frame parameter r, and the Lagrange multiplier λ_{SAD} for the SAD distortion measure. The rate term R represents the motion information and the number of bits associated with choosing the reference picture r. The rate is estimated by table-lookup using the universal variable length code (UVLC) table, even if the arithmetic entropy coding method is used. For integer-pixel search, SAD is the summed absolute difference between the original luminance signal and the motion-compensated luminance signal. In the sub-pixel refinement search, the Hadamard transform of the difference between the original luminance signal and the motion-compensated luminance signal is calculated and SAD is the sum of the absolute transform coefficients. The Hadamard transform in the sub-pixel search reflects the performance of the integer transform on the residual signal such that the expected reconstruction quality rather than the motion-compensated prediction quality is taken into account for the refinement. This favors sub-pixel positions with residuals that are highly correlated for a given summed distortion. The Lagrangian multiplier λ_{SAD} for the SAD distortion measure is related to the Lagrangian multiplier for the SSD measure (8) by

$$\lambda_{\text{SAD}} = \sqrt{\lambda}. \tag{11}$$

Further details as well as block size issues for motion estimation are discussed in [28] and [29].

C. Rate Constrained Multihypothesis Motion Estimation

For the multihypothesis mode, the encoder utilizes rate-constrained multihypothesis motion estimation. The cost function incorporates the multihypothesis prediction error of the video signal as well as the bit-rate for two picture reference parameters, two hypotheses types, and the associated motion vectors. Rate-constrained multihypothesis motion estimation is performed by the hypothesis selection algorithm [1]. This iterative algorithm performs conditional rate-constrained motion estimation and is a computationally feasible solution to the joint estimation problem [31] which has to be solved for finding an efficient pair of hypotheses.

The iterative algorithm is initialized with the data of the best macroblock type for multiple reference prediction (initial hypothesis). The algorithm continues with:

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

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

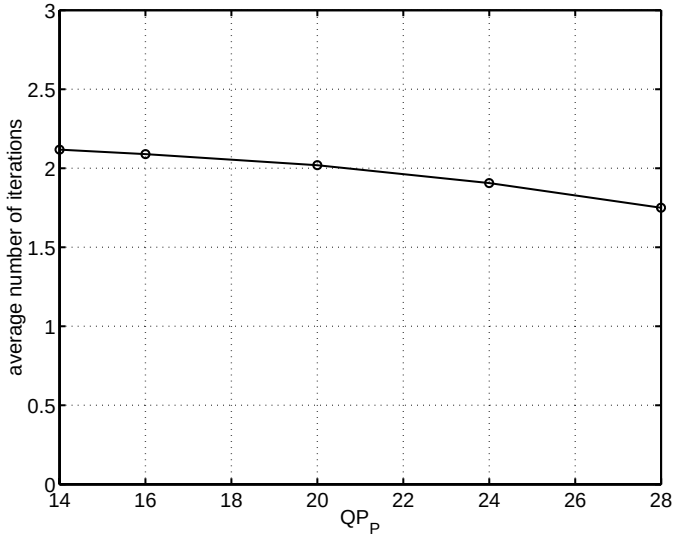


Fig. 15. Average number of iterations for multihypothesis motion estimation vs. quantization parameter for the CIF sequence *Mobile & Calendar* with 30 fps and $M = 5$ reference pictures.

Fig. 15 shows the average number of iterations for multihypothesis motion estimation with 5 reference pictures over the quantization parameter. It takes about 2 iterations to achieve a Lagrangian cost smaller than 0.5% relative to the Lagrangian cost in the previous iteration. The algorithm converges faster for higher quantization parameter values.

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

Multihypothesis prediction improves the prediction signal by allocating more bits to the side-information associated with the motion-compensating predictor. But the encoding of the prediction error and its associated bitrate 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 important tool to decide whether a macroblock should be predicted with one or two hypotheses.

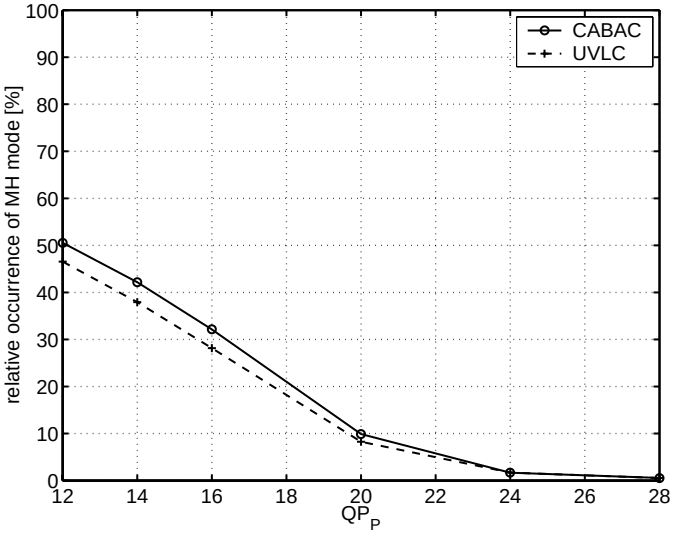


Fig. 16. Relative occurrence of the multihypothesis mode in B pictures vs. quantization parameter for the CIF sequence *Mobile & Calendar* with 30 fps. 5 past and 3 future reference pictures are used. $QP_B = QP_P + 2$.

Fig. 16 shows the relative occurrence of the multihypothesis mode in generalized B pictures over the quantization parameter for the CIF sequence *Mobile & Calendar*. 5 past and 3 future reference pictures are used. Results for both entropy coding schemes are plotted. For high bitrates (small quantization parameters), the multihypothesis mode exceeds a relative occurrence of 50 % among all B picture coding modes. For low bit-rates (large quantization parameters), the multihypothesis mode is selected infrequently and, consequently, the improvement in coding efficiency is very small. In addition, the relative occurrence is slightly larger for the CABAC entropy coding scheme since the more efficient CABAC scheme somewhat relieves the rate constraint imposed on the side-information.

D. Improving Overall Rate-Distortion Performance

When B pictures establish an enhancement layer in a scalable representation, they are predicted from reference pictures that are provided by the base layer. [2] Consequently, the quality of the base layer influences the rate-distortion trade-off for B pictures in the enhancement layer. Experimental results show that the relationship between quantization and Lagrange parameter for mode decision $\lambda = f(QP)$ should be adapted. The following experimental results are obtained with the test model software TML-9, i.e., with bi-directional prediction and independent estimation of forward and backward prediction parameters.

[2]The RPSL scheme in H.264/AVC permits the feature that B pictures can be referenced for B picture coding but not for P picture coding. This ensures that the base layer with P pictures is independent of the enhancement layer with B pictures. Note, that we do not use this feature in this section.

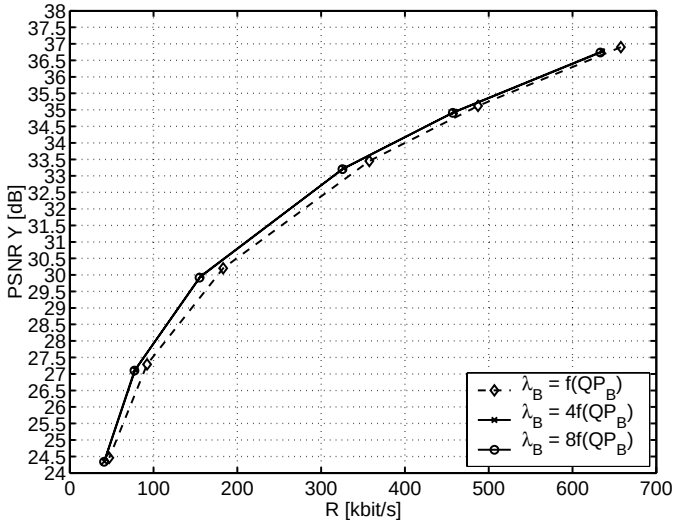


Fig. 17. PSNR of the luminance signal vs. overall bit-rate for the QCIF sequence *Mobile & Calendar* with 30 fps. Two B pictures are inserted and the influence of the $\lambda_B - QP_B$ relationship on the overall compression efficiency is investigated.

Fig. 17 shows the PSNR of the luminance signal vs. overall bit-rate for the QCIF sequence *Mobile & Calendar* with 30 fps. Three different $\lambda - QP$ dependencies are depicted. The worst compression efficiency is obtained with $\lambda_B = f(QP_B)$. The cases $\lambda_B = 4f(QP_B)$ and $\lambda_B = 8f(QP_B)$ demonstrate similar but superior efficiency for low bit-rates. The scaling of the dependency alters the bit-rate penalty for all B picture coding modes such that the overall compression efficiency is improved. The factor 4 is suggested in the test model TML-9 description.

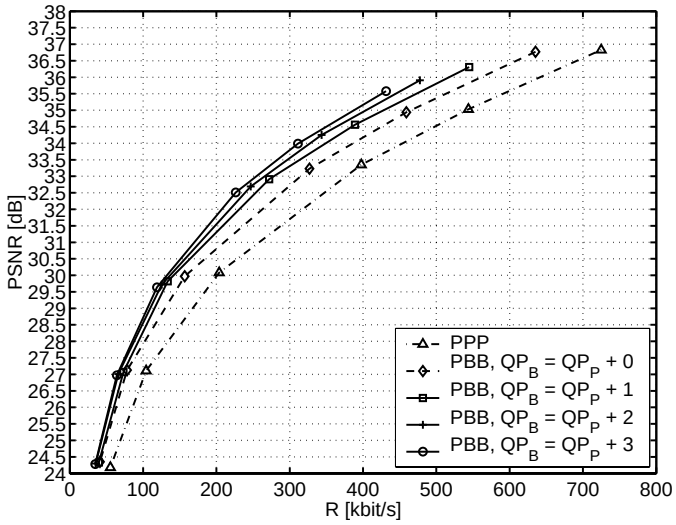


Fig. 18. PSNR of the luminance signal vs. overall bit-rate for the QCIF sequence *Mobile & Calendar* with 30 fps. Two B pictures are inserted and the influence of the B picture quantization parameter QP_B on the overall compression efficiency is investigated for $\lambda_B = 4f(QP_B)$.

Further experiments show that not only the relationship between quantization and Lagrange parameter for mode decision has to be adapted for B pictures but also the PSNR of the enhancement layer should be lowered in comparison to the base layer to improve overall compression efficiency [32]. Fig. 18 depicts also the PSNR of the luminance signal vs. overall bit-rate for the QCIF sequence *Mobile & Calendar* with 30 fps. The plot compares the compression efficiency of various layered bit-streams with two inserted B pictures. The quantization parameters of inter and B pictures differ by a constant offset. For comparison, the efficiency of the single layer bit-stream is provided. Increasing the quantization parameter for B pictures, that is, lowering their relative PSNR, improves the overall compression efficiency of the sequence.

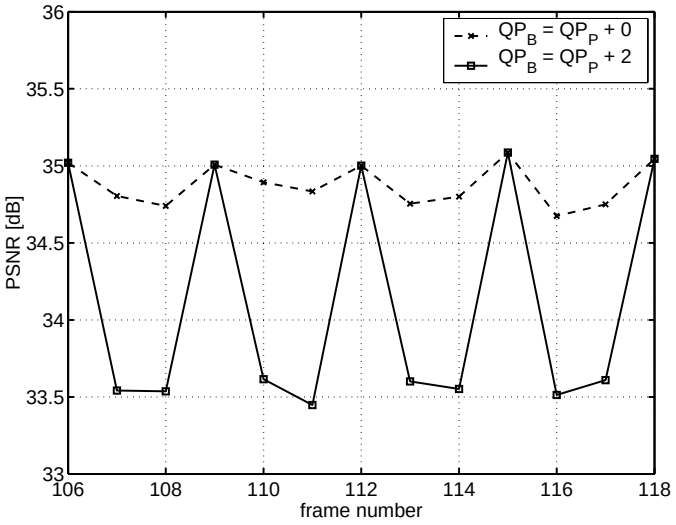


Fig. 19. PSNR of the luminance signal for individual pictures. Two B pictures are inserted. The B picture quantization parameter QP_B is incremented by 2 and the B picture Lagrange parameter $\lambda_B = 4f(QP_B)$. $QP_P = 14$.

Fig. 19 shows the PSNR of the luminance signal for individual pictures of the sequence *Mobile & Calendar* encoded with $QP_P = 14$. The PSNR of the B pictures encoded with an increment of 2 is significantly lower compared to the case with identical quantization parameter in both layers. The compression efficiency of the sequence increases by lowering the relative PSNR of the enhancement layer. For the investigated sequence, the average PSNR efficiency increases by almost 1 dB (see Fig. 18), whereas the PSNR of individual B pictures drops by more than 1 dB. In this case, higher average PSNR with temporal fluctuations is compared to lower average PSNR with less fluctuations for a given bit-rate.

Fig. 20 shows the PSNR of the luminance signal vs. overall bit-rate for the QCIF sequence *Foreman* with 30 fps. The depicted results demonstrate that not adapting the quantization parameter and the $\lambda - QP$ dependency for B pictures causes a degradation in compression efficiency if two inter pictures are replaced by B pictures, whereas the adaptation improves the PSNR by about 0.5 dB for a given bit-rate.

Depending on the video sequence, this temporal fluctuations may affect the subjective quality under playback. However, if B pictures are used as reference for B picture

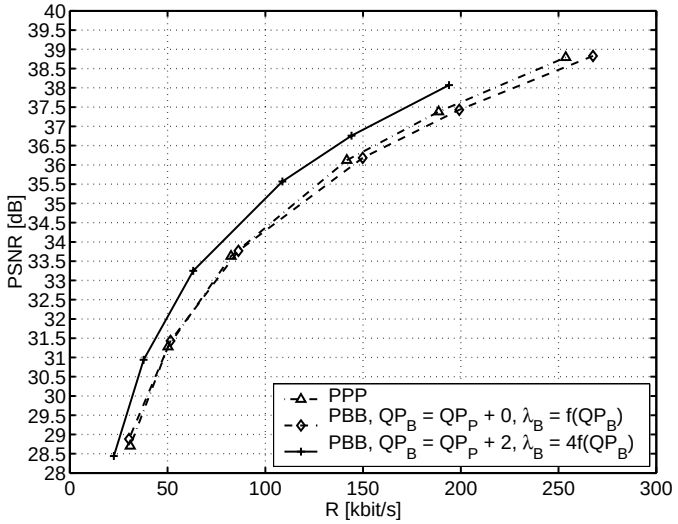


Fig. 20. PSNR of the luminance signal vs. overall bit-rate for the QCIF sequence *Foreman* with 30 fps. When replacing two inter pictures by B pictures, the values QP_B and λ_B have to be adapted for best compression efficiency. Keeping the inter picture values may lower the efficiency.

coding, the compression efficiency of B pictures will also be enhanced. In this case, a quality degradation for the B pictures is not advisable as high quality reference pictures improve prediction performance.

V. Conclusions

This paper discusses B pictures in the draft H.264/AVC video compression standard and uses the test model TML-9 to obtain experimental results. Additionally, it differentiates between picture reference selection and linearly combined prediction signals. This distinction is reflected by the term *Generalized B Pictures*. The feature of reference picture selection has been improved significantly when compared to existing video compression standards. But with respect to combined prediction signals, the draft H.264/AVC video compression standard provides new features.

The current draft specifies explicitly macroblock modes for B pictures. But a desirable definition of a generalized picture type should provide generic macroblock modes independent of the utilized reference pictures. In any case, this generalized picture type should permit linearly combined prediction signals. With this definition, generalized B pictures utilize the direct and the multihypothesis mode, whereas classic inter pictures replace the direct mode by the copy mode and disable the multihypothesis mode.

Not only in the case of B pictures, the emerging standard has improved significantly in many aspects of its design and clearly outperforms existing video compression standards.

References

[1] 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, vol. 4067, pp. 238–249.

[2] M. Flierl, T. Wiegand, and B. Girod, "Rate-constrained multihypothesis motion-compensated prediction for video coding", in *Proceedings of the IEEE International Conference on Image Processing*, Vancouver, Canada, Sept. 2000, vol. 3, 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] 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.

[5] M. Hannuksela, "Prediction from temporally subsequent pictures", Document Q15-K38, ITU-T Video Coding Experts Group, Aug. 2000, http:// standards. pictel. com/ ftp/ video-site/ 0008_Por/ q15k38.doc.

[6] ITU-T Video Coding Experts Group and ISO/IEC Moving Picture Experts Group, *Study of Final Committee Draft of Joint Video Specification (ITU-T Rec. H.264, ISO/IEC 14496-10 AVC)*, Mar. 2003, ftp:// ftp.imtc-files.org/ jvt-experts/ 2003_03_Pattaya/ JVT-G050d4.zip.

[7] M. Flierl and B. Girod, "Further investigation of multihypothesis motion pictures", Document VCEG-M40, ITU-T Video Coding Experts Group, Apr. 2001, http:// standards. pictel. com/ ftp/ video-site/ 0104_Aus/ VCEG-M40.doc.

[8] M. Flierl and B. Girod, "Multihypothesis predition for B frames", Document VCEG-N40, ITU-T Video Coding Experts Group, Sept. 2001, http:// standards. pictel. com/ ftp/ video-site/ 0109_San/ VCEG-N40.doc.

[9] ISO/IEC, *13818-2 Information Technology - Generic Coding of Moving Pictures and Associated Audio Information: Video (MPEG-2)*, 1996.

[10] W.B. Pennebaker, J.L. Mitchell, D. Le Gall, and C. Fogg, *MPEG Video Compression Standard*, Kluwer Academic Publishers, Boston, 1996.

[11] H.G. Musmann, P. Pirsch, and H.J. Grallert, "Advances in picture coding", *Proceedings of the IEEE*, vol. 73, no. 4, pp. 523–548, Apr. 1985.

[12] S. Ericsson, "Fixed and adaptive predictors for hybrid predictive/transform coding", *IEEE Transactions on Communications*, vol. 33, no. 12, pp. 1291–1302, Dec. 1985.

[13] A. Puri, R. Aravind, B.G. Haskell, and R. Leonardi, "Video coding with motion-compensated interpolation for CD-ROM applications", *Signal Processing: Image Communication*, vol. 2, no. 2, pp. 127–144, Aug. 1990.

[14] 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.

[15] 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.

[16] ITU-T, *Recommendation H.263++ (Video Coding for Low Bitrate Communication)*, 2000.

[17] ITU-T Video Coding Experts Group, *H.26L Test Model Long Term Number 9, TML-9*, Dec. 2001, http:// standards. pictel. com/ ftp/ video-site/ h26L/ tml9.doc.

[18] T. Yang, K. Liang, C. Huang, and K. Huber, "Temporal scalability in H.26L", Document Q15-J45, ITU-T Video Coding Experts Group, May 2000, http:// standards. pictel. com/ ftp/ video-site/ 0005_Osa/ q15j45.doc.

[19] K. Lillevold, "B pictures in H.26L", Document Q15-I08, ITU-T Video Coding Experts Group, Oct. 1999, http:// standards. pictel. com/ ftp/ video-site/ 9910_Red/ q15i08.doc.

[20] S. Kondo, S. Kadono, and M. Schlockermann, "New prediction method to improve B-picture coding efficiency", Document VCEG-O26, ITU-T Video Coding Experts Group, Dec. 2001, http:// standards. pictel. com/ ftp/ video-site/ 0112_Pat/ VCEG-O26.doc.

[21] K. Lillevold, "Improved direct mode for B pictures in TML", Document Q15-K44, ITU-T Video Coding Experts Group, Aug. 2000, http:// standards. pictel. com/ ftp/ video-site/0008_Por/ q15k44.doc.

[22] M. Flierl, T. Wiegand, and B. Girod, "Multihypothesis pictures for H.26L", in *Proceedings of the IEEE International Conference on Image Processing*, Thessaloniki, Greece, Oct. 2001, vol. 3, pp. 526–529.

[23] D. Marpe, G. Blättermann, and T. Wiegand, "Adaptive codes for H.26L", Document VCEG-L13, ITU-T Video Coding Experts Group, Jan. 2001, http:// standards. pictel. com/ ftp/ video-site/ 0101_Eib/ VCEG-L13.doc.

[24] D. Marpe, G. Blättermann, G. Heising, and T. Wiegand, "Further results for CABAC entropy coding scheme", Document VCEG-M59, ITU-T Video Coding Experts Group, Apr. 2001, http:// standards. pictel. com/ ftp/ video-site/ 0104_Aus/ VCEG-M59.doc.
[25] T. Stockhammer and T. Oelbaum, "Coding results for CABAC entropy coding scheme", Document VCEG-M54, ITU-T Video Coding Experts Group, Apr. 2001, http:// standards. pictel. com/ ftp/ video-site/ 0104_Aus/ VCEG-M54.doc.
[26] D. Marpe, "Context-adaptive binary coding for H.264/AVC", *IEEE Transactions on Circuits and Systems for Video Technology*, 2003, Special Issue on H.264/AVC.
[27] B. Jeon and Y. Park, "Mode decision for B pictures in TML-5", Document VCEG-L10, ITU-T Video Coding Experts Group, Jan. 2001, http:// standards. pictel. com/ ftp/ video-site/ 0101_Eib/ VCEG-L10.doc.
[28] G.J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression", *IEEE Signal Processing Magazine*, vol. 15, pp. 74–90, Nov. 1998.
[29] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control", in *Proceedings of the IEEE International Conference on Image Processing*, Thessaloniki, Greece, Oct. 2001, vol. 3, pp. 542–545.
[30] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G.J. Sullivan, "Rate-constrained coder control and comparison of video coding standards", *IEEE Transactions on Circuits and Systems for Video Technology*, 2003, Special Issue on H.264/AVC.
[31] S.-W. Wu and A. Gersho, "Joint estimation of forward and backward motion vectors for interpolative prediction of video", *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 684–687, Sept. 1994.
[32] H. Schwarz and T. Wiegand, "An improved H.26L coder using lagrangian coder control", Document VCEG-HHI, ITU-T Video Coding Experts Group, May 2001, http:// standards. pictel. com/ ftp/ video-site/ 0105_Por/ HHI-RDOpt.doc.

Markus Flierl (S'01) received the Dipl.-Ing. degree in electrical engineering from the University of Erlangen-Nuremberg, Germany, in 1997. From 1999 to 2001, he was a scholar with the Graduate Research Center at the University of Erlangen-Nuremberg. Until December 2002, he visited the Information Systems Laboratory at Stanford University, Stanford, CA. He contributed to the ITU-T Video Coding Experts Group standardization efforts. His current research interests are data compression, signal processing, and motion in image sequences.

Bernd Girod (M'80–SM'97–F'98) is Professor of Electrical Engineering in the Information Systems Laboratory of Stanford University, California. He also holds a courtesy appointment with the Stanford Department of Computer Science. His research interests include networked multimedia systems, video signal compression, and 3-d image analysis and synthesis. He received his M.S. degree in Electrical Engineering from Georgia Institute of Technology, in 1980 and his Doctoral degree "with highest honours" from University of Hannover, Germany, in 1987. Until 1987 he was a member of the research staff at the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover, working on moving image coding, human visual perception, and information theory. In 1988, he joined Massachusetts Institute of Technology, Cambridge, MA, USA, first as a Visiting Scientist with the Research Laboratory of Electronics, then as an Assistant Professor of Media Technology at the Media Laboratory. From 1990 to 1993, he was Professor of Computer Graphics and Technical Director of the Academy of Media Arts in Cologne, Germany, jointly appointed with the Computer Science Section of Cologne University. He was a Visiting Adjunct Professor with the Digital Signal Processing Group at Georgia Institute of Technology, Atlanta, GA, USA, in 1993. From 1993 until 1999, he was Chaired Professor of Electrical Engineering/Telecommunications at University of Erlangen-Nuremberg, Germany, and the Head of the Telecommunications Institute I, co-directing the Telecommunications Laboratory. He has served as the Chairman of the Electrical Engineering Department from 1995 to 1997, and as Director of the Center of Excellence "3-D Image Analysis and Synthesis" from 1995-1999. He has been a Visiting Professor with the Information Systems Laboratory of Stanford University, Stanford, CA, during the 1997/98 academic year. As an entrepreneur, he has worked successfully with several start-up ventures as founder, investor, director, or advisor. Most notably, he has been a founder and Chief Scientist of Vivo Software, Inc., Waltham, MA (1993-98); after Vivo's acquisition, since 1998, Chief Scientist of RealNetworks, Inc. (Nasdaq: RNWK); and, since 1996, an outside Director of 8x8, Inc. (Nasdaq: EGHT). He has authored or co-authored one major text-book and over 200 book chapters, journal articles and conference papers in his field, and he holds about 20 international patents. He has served as on the Editorial Boards or as Associate Editor for several journals in his field, and is currently Area Editor for the IEEE Transactions on Communications as well as member of the Editorial Boards of the journals EURASIP Signal Processing, the IEEE Signal Processing Magazine, and the ACM Mobile Computing and Communication Review. He has chaired the 1990 SPIE conference on "Sensing and Reconstruction of Three-Dimensional Objects and Scenes" in Santa Clara, California, and the German Multimedia Conferences in Munich in 1993 and 1994, and has served as Tutorial Chair of ICASSP-97 in Munich and as General Chair of the 1998 IEEE Image and Multidimensional Signal Processing Workshop in Alpbach, Austria. He has been the Tutorial Chair of ICIP-2000 in Vancouver and the General Chair of the Visual Communication and Image Processing Conference (VCIP) in San Jose, CA, in 2001. He has been a member of the IEEE Image and Multidimensional Signal Processing Committee from 1989 to 1997 and was elected Fellow of the IEEE in 1998 'for his contributions to the theory and practice of video communications.' He has been named 'Distinguished Lecturer' for the year 2002 by the IEEE Signal Processing Society.