GENERALIZED B PICTURES

Markus Flierl

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

Bernd Girod

Information Systems Laboratory
Stanford University
Stanford, CA
bgirod@stanford.edu

ABSTRACT

This paper discusses generalized B pictures in the context of the emerging JVT/H.26L 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 can further improve the emerging JVT/H.26L compression standard.

1. 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 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 under consideration for the emerging video compression standard. It is desirable that an arbitrary pair of reference pictures can be signaled to the decoder [6, 7]. 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 [8, 9] is achieved, where top and bottom fields are averaged to form the final prediction.

2. BI-DIRECTIONAL VS. MULTIHYPOTHESIS MODE

In the following, we will outline the difference between the bi-directional macroblock mode, which is specified in the test model TML-9 [10], and the multihypothesis mode proposed in [7] and discussed in the previous section. A bi-directional prediction type only allows a linear combination of a forward / backward prediction pair (Fig. 1). The draft 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 [11]. 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.

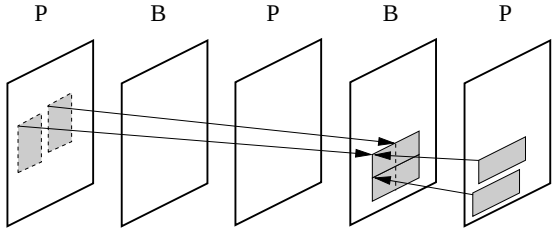

Figure 1: A bi-directional prediction mode allows a linear combination of one past and one subsequent macroblock prediction signal. The inter pictures are denoted by P.

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 (Fig. 2). 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.

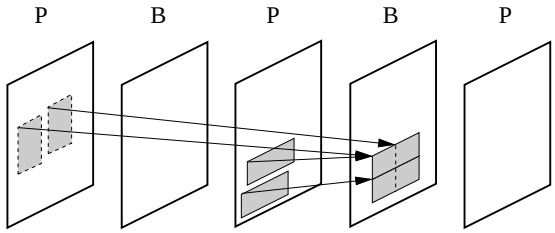

Figure 2: The multihypothesis mode also allows a linear combination of two past macroblock prediction signals.

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 test model TML-9 [10]. 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.

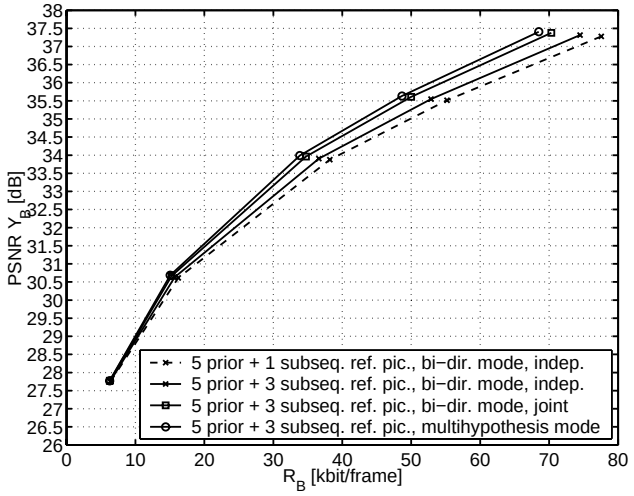

Figure 3: PSNR of the B picture luminance signal vs. B picture bit-rate for the CIF sequence *Mobile & Calendar* with 30 fps. Two B pictures are inserted after each inter picture. $QP_B = QP_P$. The multihypothesis mode is compared to the bi-directional mode.

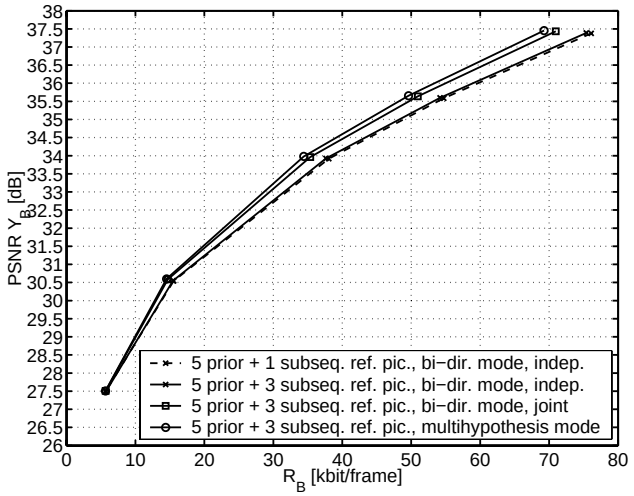

Figure 4: PSNR of the B picture luminance signal vs. B picture bit-rate for the CIF sequence *Flowergarden* with 30 fps. Two B pictures are inserted after each inter picture. $QP_B = QP_P$. The multihypothesis mode is compared to the bi-directional mode.

Figs. 3 and 4 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.

3. 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 [6]. 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.

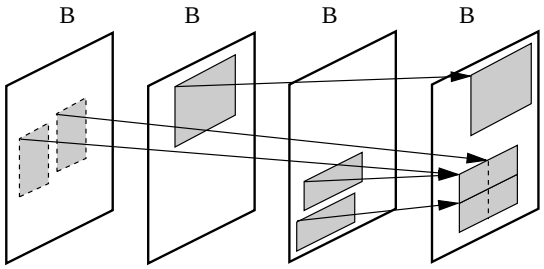

Figure 5: Generalized B pictures with forward-only prediction utilize multiple reference picture prediction and multihypothesis motion-compensated prediction. The multihypothesis mode uses two hypotheses chosen from past reference pictures.

Fig. 5 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 [10] allows seven different block sizes which will be the seven hypotheses types in the multihypothesis mode. The TML-9 draft allows for inter modes only one picture reference parameter per macroblock and assumes that all sub-blocks can be found on that specified reference picture. This is different from the H.263 standard, where multiple reference picture prediction utilizes picture reference parameters for both macroblocks and 8×8 blocks [4].

We investigate the rate-distortion performance of generalized B pictures with forward-only prediction and compare them to TML-9 inter pictures for various numbers of reference pictures. Fig. 6 shows the bit-rate values at 35 dB PSNR of the luminance signal over the number of reference pictures M for the CIF se-

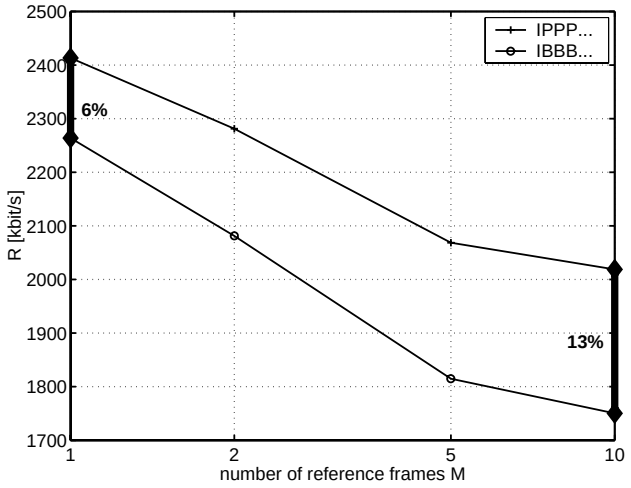

Figure 6: Average bit-rate at 35 dB PSNR vs. number of reference pictures for the CIF sequence *Mobile & Calendar* with 30 fps. Generalized B pictures with forward-only prediction are compared to inter pictures.

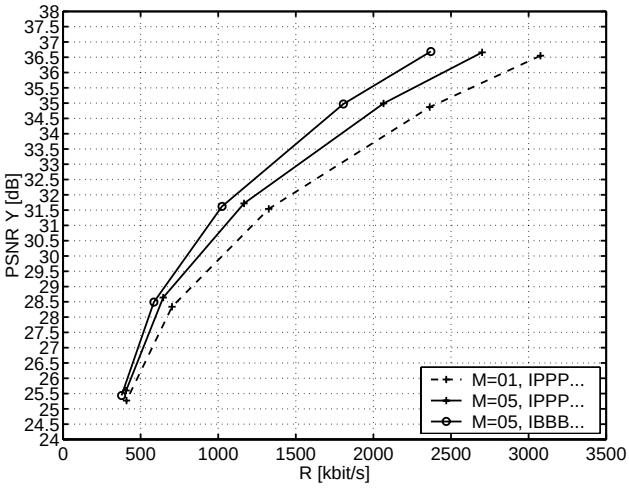

Figure 7: PSNR of the luminance signal vs. overall bit-rate for the CIF sequence *Mobile & Calendar* with 30 fps. Generalized B pictures with forward-only prediction are compared to inter pictures.

quences *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 TML-9 inter pictures (IPPP...) and the generalized B pictures with forward-only prediction (IBBB...) is shown.

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 [12]. 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 [12]. This observation is independent of the implemented multihypothesis prediction scheme [13].

Fig. 7 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.

4. 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) [14, 15]. 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 [15]. 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.

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

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

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

$$e_k(C) = |mvd_k(\text{left})| + |mvd_k(\text{top})|. \tag{2}$$

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

$$e_k(C) = |mvd_k(\text{first hypothesis})|. \tag{3}$$

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.

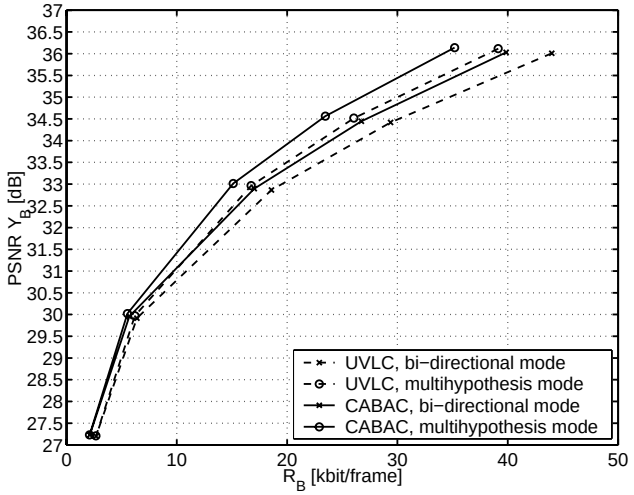

Figure 8: PSNR of the B picture luminance signal vs. B picture bitrate for the CIF sequence *Flowergarden* with 30 fps. Two B pictures are inserted after each inter picture. 5 past and 3 future inter pictures are used for predicting each B picture. $QP_B = QP_P + 2$ and $\lambda_B = 4f(QP_P)$. The multihypothesis mode and the bidirectional mode with independent estimation are compared for both entropy coding schemes.

Fig. 8 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 bidirectional 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.

5. CONCLUSIONS

This paper discusses B pictures in the emerging JVT/H.26L compression standard [16]. 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 joint working draft can further be refined.

6. 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 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] 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] 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.

[7] 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.

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

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] ITU-T Video Coding Experts Group and ISO/IEC Moving Picture Experts Group, *Working Draft Number 2, Revision 7*, Apr. 2002, ftp:// ftp.imtc-files.org/ c:/ inetpub/ ftpsites/ imtc/ jvt-experts/ draft_standard/ jwd2r7.zip.