NEW BI-PREDICTION TECHNIQUES FOR B PICTURES CODING

Xiangyang Ji[1], Debin Zhao[2], Wen Gao[1], Qingmin Huang[1], Siwei Ma[1], Yan Lu[2]

Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, China[1]
Department of Computing Science, Harbin Institute of Technology, Harbin, 150001, China[2]
{xyji, dbzhao, wgao, qmhuang, swma, ylu}@jdl.ac.cn

ABSTRACT

In the conventional video codec, B pictures usually employ five prediction modes, including bi-prediction, direct prediction, forward prediction, backward prediction and intra modes. Among them, the direct and bi-prediction modes, whose prediction values are achieved from both forward and backward reference pictures, are more efficient in exploiting the temporal correlation between the reference pictures and the current B picture. In addition, the direct mode does not require any bits for coding the motion vectors. Therefore, the blocks coded with direct or bi-predictive mode are usually much more than the other modes. To further take advantage of the bi-prediction, in this paper we propose a new bi-predictive coding technique, which can achieve the good tradeoff between the bit-rate saving for motion vector coding and the prediction accuracy. Moreover, we also propose the spatial motion vector prediction and motion vector scaling techniques to improve the accuracy of derived direct mode motion vectors. All these techniques have been adopted in AVS[1] video coding standard.

1. INTRODUCTION

In hybrid video coding, B pictures use both future and past pictures as references. The prediction signals are obtained by linearly combining the prediction values separately from forward reference picture and backward reference picture based on motion compensation. B pictures can achieve higher compression ratio by more effectively exploiting the correlation between reference pictures and current B picture, especially for copping with occlusion, uncovering problem caused by zooming, non-linear motion and so on. B pictures also can be coarsely quantized in some application because a B picture is not used as a reference picture and does not propagate errors. Therefore, the B pictures play an important role in hybrid video coding in term of both coding performance and transmission efficiency, in particular in some real-time video transmission services.

In the early video coding systems, considering the real-time coding and solving the overflow of bit rate buffer, some pictures will be skipped or dropped at the encoder. At the decoder, these pictures (B pictures) will be reconstructed from transmitted pictures (I/P pictures). Such a technique is referred to as frame interpolation. Therefore, there are no residual and motion vectors transmitted for interpolated frames. The motion vectors of every block in B pictures are derived from the transmitted pictures.

However, frame interpolation is only suitable for video sequences with simple translational motion. For non-linear motion, especially for uncovered region, frame interpolation technique cannot achieve the acceptable coding performance. Therefore, a new bi-predictive coding technique [3] that needs to code a pair of forward and backward motion vectors has been proposed to MPEG-1 standard to replace conventional frame interpolation technique. The bi-predictive coding technique can efficiently solve the problems presented in the sequence with non-linear motion rather than the traditional frame interpolation technique.

In addition, both MPEG-4 [1] and H.264/AVC standards [2] have adopted the direct mode, which is originated from the aforementioned frame interpolation technique. As illustrated in Fig.1, direct mode takes advantage of bi-directional prediction and does not need to transmit motion vectors by deriving its forward and backward motion vectors from motion vector used in its co-located block [4]. The forward and backward direct mode motion vectors are calculated as follows:

$$MV_F = \frac{tb}{td} \times MV_D \tag{1}$$

$$MV_B = \frac{tb - td}{td} \times MV_D \tag{2}$$

where *tb* denotes the temporal distance between the current B picture and the forward reference picture, *td* denotes the temporal distance between the forward reference picture and the backward reference picture, and MV_D denotes the motion vector of the co-located block in the backward reference picture.

However, for direct mode, the motion vectors of some co-located blocks in the backward reference picture cannot be obtained in some cases, e.g. uncovered boundaries, overlapped blocks, the luminance change in adjacent pictures, etc, which may lead to many blocks coded with intra mode. If we fail in the derivation of motion vector and only set it zero, the accuracy of the derived motion vectors would be deteriorated. In addition, for existing bi-prediction mode, both the forward and backward motion vectors need to be coded. However, in some cases, especially for sequences with irregular motion, the cost of coding motion vector accounts for high portion.

To tackle these problems, this paper presents a new bi-prediction mode, i.e. symmetric mode, to efficiently save the bits for coding motion vectors compared with conventional bi-prediction mode. Furthermore, both spatial motion vector prediction and motion vector scaling techniques for direct mode

[1] The China audio-video coding standard (AVS) targeting at higher coding efficiency and lower complexity for high definition video coding than the existing video standards

0-7803-8603-5/04/$20.00 ©2004 IEEE.

are proposed to improve the accuracy of derived direct mode motion vectors. Spatial motion vector prediction for direct mode is not a new idea. However, it is implemented with a new method rather than the existing H.26x and MPEG standards, and so it is still introduced in this paper. In the following section, we are focused on presenting the proposed method and then the simulated results are presented.

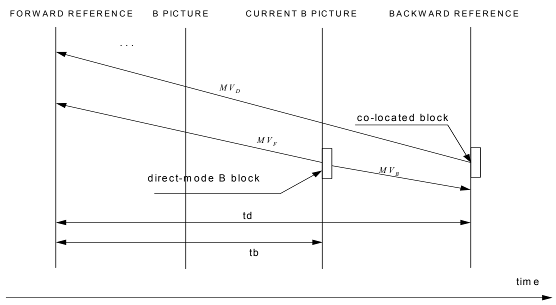

Fig.1. the direct mode block with forward and backward motion vectors MV_F, MV_B derived from motion vector MV_D of co-located block

2. BI-PREDICTION FOR B PICTURES

2.1 Symmetric Mode

As discussed previously, since existing bi-predictive mode need to code a pair of forward and backward motion vectors for each block, symmetric mode with single coded motion vector can efficiently save the bits coding motion vectors as a new bi-predictive mode. The features of this mode are described as follows:

1) Only coding single motion vector (forward motion vector) for each block is required. Motion vector predictor for differential motion vector coding is derived from spatially adjacent blocks, namely three previous coded motion vectors from the left, the above, and the above-right blocks (or the above-left block if motion vector of the right block is not available). Because we just code single motion vector with forward direction for each block, the motion vectors of spatially adjacent blocks with backward prediction mode are not available for this motion vector predictor.

2) As shown in Fig.2, the backward motion vector of the current block can be derived from forward motion vector according to the temporal distance between forward and backward reference pictures and the current B picture:

$$MV_B = \frac{tp}{tb} \times MV_F \qquad (3)$$

where tb denotes the temporal distance between the current B picture and the forward reference picture, tp denotes the temporal distance between the current B picture and the backward reference picture, and MV_F denotes the forward motion vector of the current block pointing to the forward reference picture, MV_B denotes the backward motion vector of the current block pointing to the backward reference picture.

3) For bi-prediction picture, although the coding performance can be improved by joint estimation, independently estimating forward and backward prediction block is practical because joint estimation of forward and backward motion vectors is computationally too expensive to be implemented in reality. However, for symmetric mode, joint estimation algorithm can be easily implemented with the same process as conventional forward or backward independent search solution. Meanwhile, to further simplify the search process, we can make such joint estimation just limit to fractional sample with less search range. Accordingly, symmetric mode can efficiently not only save the bits coding motion vectors for bi-prediction mode but also further exploit the temporal correlation between the reference pictures and the current B picture.

In fact, to better exploit the temporal correlation for different cases, the symmetric mode with single motion vector transmitted and existing bi-prediction mode with a pair of forward and backward motion vectors transmitted can be combined by introducing one flag at picture level. In coding process, such combination can be implemented by two-pass coding or pre-processing for every B picture. The better one of them at picture level can be selected according to Rate Distortion Optimization decision:

$$J = D + \lambda \times R$$

Where $\lambda = 0.85 \times 2^{((QP-12)/3)}$, is a Lagrangian multiplier varied with QP value, D is the distortion of each B picture and R is the total bitrate of each B picture .

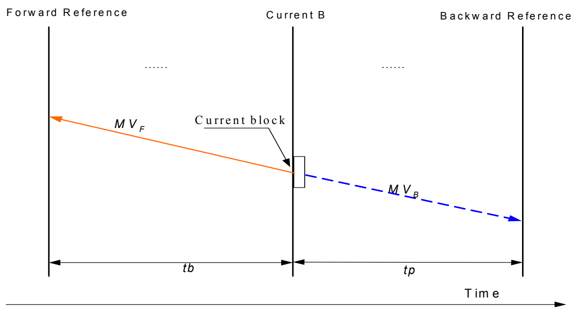

Fig.2. the backward motion vector of current block is derived from forward motion vector in symmetric mode

2.2 Direct Mode

A. Spatial prediction technique

In H.264/AVC, a new type of direct mode, i.e. Spatial Direct Mode [5], is also utilized, wherein the motion vectors are derived from the motion vectors of spatially adjacent blocks. The spatial motion vector predictor is the median of the three previous coded motion vectors from the left, the above, and the above-right blocks (or the above-left blocks if motion vector of the right block is not available), as shown in Fig.3. Such spatial prediction technique can efficiently exploit the spatial correlation. At the same time, if there are no available motion vectors for one direction, Spatial Direct Mode will only use the single prediction with available motion vectors to efficiently tackle scene change etc. in video sequences.

In H.264/AVC, when the co-located block in the backward reference picture is coded with intra mode, the corresponding motion vectors for scaling will be set as zero. However, this method is inefficient in most cases. To tackle this problem, AVS introduced the spatial prediction scheme in direct mode. That is to say, direct mode can further exploit the spatial redundancy by Spatial Direct Mode technique when the co-located block is coded in intra mode.

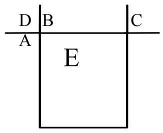

Fig.3. median prediction of motion vectors.

B. Scaling motion vectors for direct mode

In conventional video coding standard, the maximum decoder buffers for reference pictures depend on the total reference picture number of B picture. For example, in MPEG-2 and MPEG-4 decoder, B pictures need two reference pictures buffer and P picture only need one reference picture buffer. Therefore, for P picture decoding, one of reference picture buffers is never used. In AVS standard, for P picture coding, up to two reference pictures are used. By this means, we can efficiently use the reference pictures buffer to improve the coding performance and at the same time no extra buffer is required.

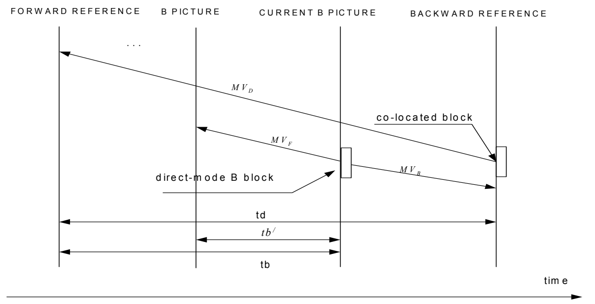

Fig.4. Scaling motion vector for direct mode

However, in the case using the fixed size reference pictures buffer, the total forward reference picture number of P picture will be more than the total forward reference picture number of B pictures. As illustrated in Fig.4, to guarantee the forward reference picture of B picture available, if the co-located block in backward reference picture points to corresponding reference picture which will exceed the maximum forward reference picture that the block in B picture can point to, the forward reference picture for current B block should be set to the maximum forward reference picture of B picture. Moreover, to get better forward motion vector for direct mode, we should introduce a new motion vector scaling technique to get better prediction values. The general formula, which is suitable for both the normal and above cases, for deriving forward and backward direct mode motion vectors, is as follows:

$$MV_{\mathrm{F}} = \frac{tb^{/}}{td} \times MV_{D} \quad (4)$$

$$MV_{\mathrm{B}} = \frac{tb - td}{td} \times MV_{D} \quad (5)$$

where $tb^{'}$ denotes the temporal distance between the current picture and the maximum forward reference picture, td denotes the temporal distance between the forward picture, i.e. the reference picture used by the co-located block in backward reference picture, and the backward reference picture, and MV_F denotes the forward motion vector of the current block pointing to the forward reference picture, MV_B denotes the backward motion vector of the current block pointing to the backward reference picture.

3. SIMULATED RESULTS

To evaluate the general performance of the symmetric mode compared with existing bi-prediction mode. We integrated the proposed symmetric mode into the H.264/AVC reference software JM6.1e [6]. The test sequences include *coastguard, tempete, mobile, flower,* and *akiyo* with 30fps in the CIF format (352×288) and *spincalendar, night, crew* and *harbour* with 30fps in the HD format (1280×720). The AVS reference software RM5.0 was used to evaluate the performance of direct mode with proposed method. The test sequences include *spincalendar, night, crew, city* and *harbour* with 30fps in the HD format (1280×720). Considering AVS software RM5.0 already supported the proposed direct mode coding technique since it has been adopted by AVS standard. The experimental results achieved by RM5.0 with the proposed direct mode coding technique and without it are presented here for comparison. The test conditions are shown in Table 1. To evaluate the average PSNR vs. bitrate, we employ the method described in [7], which is widely used during H.264/AVC development. The detailed results are showed in Table 2 and Table 3.

Table 2 shows, in general, the proposed symmetric mode coding technique can efficiently improve the performance for sequences with high motion. Rate-distortion curves for mobile sequence in Fig.5 further verify such a conclusion. But for other sequences with low or accelerated motion, conventional bi-prediction tends to get better results.

Table 1. Test conditions

Reference software	H.264/AVC jm61e	AVS rm50
MV resolution	1/4 pel	1/4 pel
Hadamard	ON	ON
RD optimization	ON	ON
Search Range	±16	±48
Restrict Search Range	no restrictions	no restrictions
Reference Frames	1	2
Symbol Mode	CAVLC	CAVLC
GOP structure	IBBPBBP…	IBBPBBP…
QP	20,24,28,32	27,30,35,40
Loop Filter	ON	ON

Table 3 shows the proposed direct mode coding technique can efficiently improve the coding performance. For *spincalendar* sequence with spin motion, the maximum PSNR gain and bitrate savings are separately up to 1.661db, 35.34% in terms of B-pictures and separately up to 0.218db, 8.86% in terms of all pictures. Fig. 6 shows the rate-distortion curves of all pictures for the test sequences of *spincalendar*. It is obvious that the proposed direct mode coding technique tends to efficiently improve coding performance in high bitrate.

4. CONCLUSIONS

In this paper, we have presented a new prediction mode for B pictures coding. The proposed symmetric mode can be taken as an alternative type of bi-directional prediction. One of the major advantages of symmetric mode coding is that it can save the bits for coding motion vectors for sequences with high motion. Since the traditional bi-prediction mode is more efficient in exploiting the temporal redundancy in many cases, it is more advantageous of combining these two techniques so as to achieve the better coding performance. In addition, for differential motion vector coding of symmetric mode, we only employ the forward motion vector predictor. Therefore, how to use both forward and backward motion vector predictor should be a research topic to be further studied.

5. ACKNOWLEDGMENTS

This research is supported by National Natural Science Foundation of China (60333020), National Hi-Tech Research Program of China (2002AA119010) and National Fundamental Research and Development Program of China (2001CCA03300).

6. REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 N2502, FDIS of ISO/IEC14496-2, "Generic coding of audio-visual objects: Part 2- Visual," (MPEG-4), Nov. 1998.

[2] 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, JVT-G050r1,ftp://ftp.imtc-files.org/jvt-experts/

[3] 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.

[4] M.Flierl and B.Girod. "Generalized B Pictures and the Draft H.264/AVC Video Compression Standard", IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 587-597, July 2003.

[5] Alexis M. Tourapis, Feng Wu, Shipeng Li, "Direct mode coding for bi-predictive pictures in the JVT standard", ISCAS2003, vol. 2, 700-703, Thailand, 2003.

[6] JVT Reference Software, Version 61e, http://bs.hhi.de/~suehring/tml/download/

[7] G. Bjontegaard, "Calculation of average PSNR differences between RD-Curves", Doc. VCEG-M33, Mar. 2001

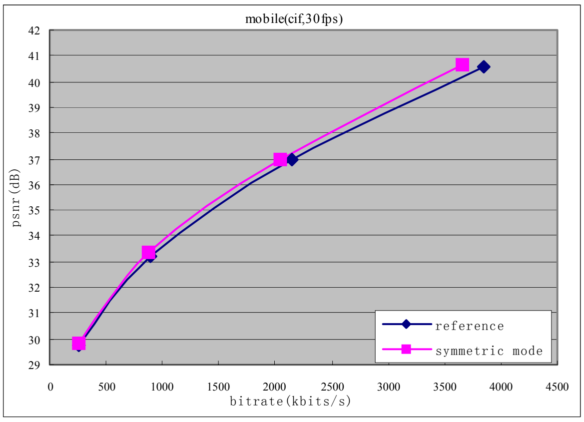

Fig.5. Rate-distortion curves for B pictures

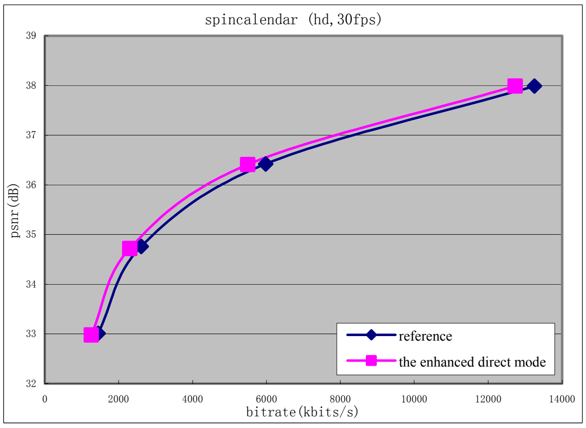

Fig.6. Rate-distortion curves for all pictures

Table 2. Performance comparisons between symmetric mode and conventional bi-predictive mode

Video sequences		spincalendar	crew	night	harbour	coastguard	mobile	tempete	flower	stefan
all pictures	*Ave PSNR* gain	0.121	0.006	0.044	-0.044	0.017	0.154	0.070	0.066	-0.058
	Ave BR Saving	3.76%	0.19%	1.20%	-1.10%	0.38%	3.18%	1.60%	1.32%	-1.61%
B pictures	*Ave PSNR gain*	0.240	0.010	0.087	-0.089	0.035	0.287	0.131	0.143	-0.118
	Ave BR Saving	7.24%	0.37%	2.31%	-2.28%	0.81%	5.73%	2.92%	2.72%	-3.36%

Table 3. Performance comparisons for direct mode with the proposed method and without it

Video sequences		spincalendar	night	harbour	crew	city
all pictures	Ave PSNR gain	0.218	0.031	0.034	0.029	0.066
	Ave BR Saving	8.86%	1.04%	1.11%	1.31%	2.44%
B pictures	Ave PSNR gain	1.661	0.085	0.098	0.079	0.269
	Ave BR Saving	35.34%	3.12%	3.27%	3.45%	7.87%