Encoding-Assisted Temporal Direct Mode Decision for B Pictures in H.264/AVC

Yan-Neng Fang and Yinyi Lin
Department of Communication Engineering
National Central University, Taiwan
Email: yilin@ce.ncu.edu.tw

Hui-Jane Hsieh
Department of Information Management
Chien Hsin University of Science and Technology, Taiwan

Abstract — **This paper proposes an encoding-assisted temporal DIRECT mode decision algorithm for H.264/AVC inter bi-predictive (B) frame video sequences to improve the coding efficiency. In the proposed algorithm, we employ motion vectors (MVs) of co-located block and its eight neighboring blocks for DIRECT mode decision. In addition, the weight selection for bidirectional prediction is also considered. The best MV and weight to minimize the sum of absolute prediction error is selected for DIRECT mode decision. The experimental results reveal that the proposed algorithm achieves average 0.1 dB PSNR gain or equivalently average 1.6% bit-rate reduction, compared to the conventional DIRECT mode coding that only uses the MV of the co-located MB for DIRECT mode decision.**

Keywords: Direct mode decision, H.264/AVC, Prediction motion vector, Rate-distortion optimization (RDO)

I. INTRODUCTION

The latest H.264/AVC achieves better performance in both PSNR and visual quality at the same bit rate, compared to prior video coding standards. This is due to that H.264/AVC features many advanced techniques, such as variable block sizes mode decision and multiple reference frames motion estimation etc., and also due to the consideration of generalized bi-predictive (B) frame video coding [1]. Another important technique is the uses of Lagrangian rate-distortion optimization (RDO).

In the H.264/AVC encoder both inter and intra mode predictions are provided in both predictive (P) and bi-predictive (B) frames. The inter mode prediction provides seven modes for inter-frame motion estimation, changing among 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, and 4x4. They are performed in each MB to achieve the best coding efficiency. The intra mode prediction offers I4x4MB prediction mode and I16x16MB prediction mode.

The inter B frame coding can use backward as well as forward frames for multiple predictions. As a result, high percentage of bits is required to encode motion information such as prediction mode, motion vector and reference frame. To alleviate high overhead problem, in addition to inter and intra modes the SKIP mode and DIRECT mode are also introduced in both P and B frames, respectively. In the SKIP or DIRECT mode coding, the motion information is obtained directly from previously encoded MB or blocks and motion information is not needed to transmit, leading to great overhead reduction within the bit stream.

The temporal DIRECT mode decision suggested in H.264/AVC is simple but not effective since in many cases the MV of co-located MB or block does not represent the true motion of the current MB or block [2][3]. This could cause severe prediction errors resulting in heavy redundant coding bits. In this paper, we propose an efficient temporal DIRECT mode decision algorithm for H.264/AVC B frame video sequences to improve its coding efficiency. In the suggested technique, in addition to the MV of the co-located block we also employ MVs of its neighboring blocks for DIRECT mode decision.

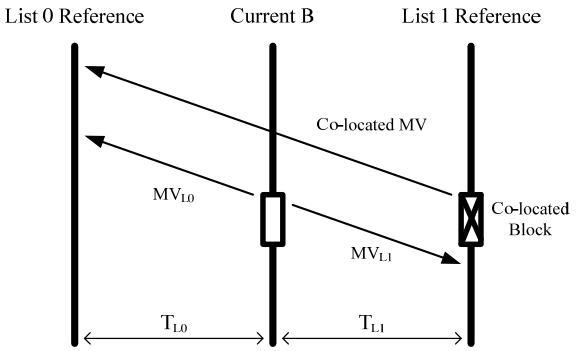

Fig. 1 Bidirectional Prediction in DIRECT mode

II. DIRECT MODE DECISION USED IN H.264/AVC

In the H.264/AVC encoder, the inter mode prediction for B frame can use forward as well as backward reference frames (namely *List 0 reference* and *List 1 reference*) for multiple predictions. The temporal DIRECT mode decision uses bidirectional predictions, and the forward MV_{L0} and backward motion vectors MV_{L1} are derived from the motion vector $MV_{co-located}$ used in the co-located block in the sub-sequential reference frame, i.e., the first *List 1 reference* frame. As illustrated in Fig. 1, the motion vectors MV_{L0} and MV_{L1} for temporal DIRECT mode blocks are calculated as

Published by Atlantis Press, Paris, France.

$$MV_{L0} = \frac{T_{L0}}{T_{L0}+T_{L1}} \cdot MV_{co-located} = (r_0, s_0) \quad (1)$$

and

$$MV_{L1} = \frac{T_{L1}}{T_{L0}+T_{L1}} \cdot MV_{co-located} = (u_0, v_0) \quad (2)$$

where T_{L0} and T_{L1} are the distances between the current frame and the forward/backward reference frames in *List 0 reference* and *List 1 reference* respectively. The bidirectional prediction for the DIRECT mode is obtained by averaging associated blocks in these references

$$\tilde{B}_n(i,j) = \frac{1}{2} B_{L0}(i+r_0, j+s_0) + \frac{1}{2} B_{L1}(i-u_0, j-v_0) \quad (3)$$

The DIRECT mode decision allows residual coding of the prediction error between the current block $B(i,j)$ and the prediction block $\tilde{B}_0(i,j)$. There are three types of DIRECT mode used in H.264/AVC based upon the residual information and the block size: DIRECT 16x16, DIRECT 8x8 and B SKIP 16x16. The residual information are transmitted in the bit-stream for both DIRECT 16x16 and DIRECT 8x8; while no residual information transmitted for B SKIP 16x16.

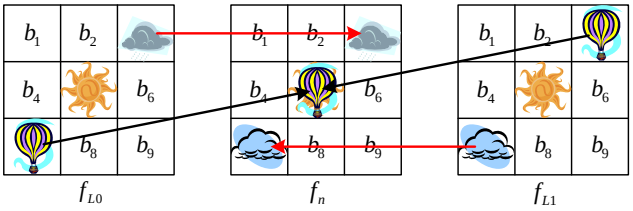

Fig. 2 Accurate prediction for DIRECT mode decision

III. ENCODING-ASSISTED DIRECT MODE DECISION FOR B PICTURES

The bidirectional prediction using (3), taking the MV of co-located block in the first *List 1 reference* frame as the estimated MV, is a simple yet efficient approach for DIRECT mode decision. The prediction error however becomes highly critical in occlusion regions or when the MV of the co-located block does not present the true motion of the current block. As illustrated in Fig. 2, when an object is moving from block b_7 in the *List 0 reference* frame f_{L0} to block b_3 in the *List 1 reference* frame f_{L1} with a constant MV, a new object comes into sight in the uncovered region and an existing object goes out of sight in the covered region in the current B frame f_n. Another object is also covered in block b_3 in the *List 1 reference* frame f_{L1}. These areas are referred to as occlusion areas. In addition, the MV of the co-located block b_5 in the *List 1 reference* frame f_{L1} (i.e., inter coded with zero MV in this case) does not represent the true MV of the current block b_5 in the current B frame f_n. As a result, the DIRECT mode decision proposed in H.264/AVC, using equal weight and MV of the co-located block b_5 in the first *List 1 reference* frame f_{L1}, cannot produce good prediction for these video occlusions, and it leads to serious prediction errors.

As illustrated in Fig. 2, $B_{L0}(i,j)$ and $B_{L1}(i,j)$ should be respectively used to predict the blocks b_3 and b_7 in the current B frame for DIRECT mode. In similar, the MV of the block b_3 in the *List 1 reference* frame should be used to predict current block b_5:

$$\tilde{B}_n(i,j) = \frac{1}{2} B_{L0}(i+r_3, j+s_3) + \frac{1}{2} B_{L1}(i-u_3, j-v_3) \quad (4)$$

where (r_3, s_3) and (u_3, v_3) are the MVs for forward and backward prediction blocks derived from the MV of the block b_3 (as denoted as MV_3) in the *List 1 reference* frame

$$MV_{L0} = \frac{T_{L0}}{T_{L0}+T_{L1}} \cdot MV_3 = (r_3, s_3) \quad (5)$$

and

$$MV_{L1} = \frac{T_{L1}}{T_{L0}+T_{L1}} \cdot MV_3 = (u_3, v_3) \quad (6)$$

To achieve more accurate prediction in DIRECT mode decision, we propose a general bidirectional prediction for DIRECT mode which is expressed as

$$\tilde{B}_n(i,j) = w_{L0} B_{L0}(i+r_k, j+s_k) + w_{L1} B_{L1}(i-u_k, j-v_k) \quad (7)$$

where w_{L0} and w_{L1} are the weights for forward and backward prediction blocks respectively in DIRECT mode with

$$w_{L0} + w_{L1} = 1 \text{ and } w_{L0}, w_{L1} \in \{0, \tfrac{1}{2}, 1\} \quad (8)$$

where $w = (w_{L0}, w_{L1}) = (0,1)$ is suitable to appearing objects (i.e., block b_7 in Fig. (2) while $w = (w_{L0}, w_{L1}) = (1,0)$ for disappearing objects (i.e., block b_3) in occlusion areas of the current B frame. For no occlusion areas, the weight

$w=(w_{L0},w_{L1})=(\frac{1}{2},\frac{1}{2})$ is the best choice.

In the proposed DIRECT mode decision, the estimations of $w=(w_f,w_b)$ and $v=(v_x,v_y)$ are accomplished in the encoder by minimizing

$$(MV,W)=\arg\min_{MV_k\in V_s}\min_{W\in W_s}\sum_{i,j}\left\|B_n(i,j)-\tilde{B}_n(i,j)\right\|_\gamma \qquad (9)$$

where γ designates the selected norm, and $\gamma=1$ is used in the experiment to measure prediction errors between the current block $B_n(i,j)$ and the bidirectional prediction block $\tilde{B}_n(i,j)$ given in (7). The parameter V_s represents the set of all MV candidates MV_k in co-located and its neighboring blocks. The prediction error can be reduced when the best MV or weight is used for DIRECT mode decision, leading to less coding bits for the redundant information.

	Prob.	Code_num	Codeword
Co-located	0.410	0	1
w=(1,0)	0.041	10	0001011
w=(0,1)	0.067	2	011
Left-Top	0.093	1	010
Left	0.066	3	00100
Left-Bottom	0.058	5	00110
Bottom	0.062	4	00101
Right-Bottom	0.058	6	00111
Right	0.054	7	0001000
Right-Top	0.047	8	0001010
Top	0.044	9	0001011

TABLE I Exp-Golomb code for extra overhead

Although the redundant coding bits can be lowered in DIRECT mode decision when the best MV or weight selected from co-located and its neighboring blocks is used, the extra overhead that indicates the MV of which block or weight values used for bidirectional prediction is required for transmission. More neighboring blocks employed for MV or weight selection, more heavy extra overhead required for transmission in DIRECT mode. The extra overhead degrades the coding performance.

To reduce performance degradation introduced in the extra overhead, in this paper we only consider the weights $w=(w_{L0},w_{L1})=(0,1)$ and $w=(w_{L0},w_{L1})=(1,0)$ for co-located block. In addition, we take into account the MVs of the co-located block as well as its eight neighboring blocks for MV selection. As a result, we need to send extra overhead describing the eleven cases when the DIRECT mode is finally determined as the best mode through RDO mode decision.

To comply with the H.264/AVC encoder, we employ the Exp-Golomb code as the entropy encoder to encode the extra overhead. The extra overhead is inserted after *mb_type* that describes the best mode for the encoding block. If the best mode is the DIRECT mode, the extra overhead describing MV and weight information is then encoded using the Exp-Golomb entropy encoder, based on the probability distribution of best MV and weight. The one with higher probability is mapped with the shorter codeword, and vice versa. An intensive experiment was conducted on many video sequences to investigate the average probability distribution of MV and weight information. The probability distribution and associated codewords for extra overhead is documented in TABLE I.

IV. EXPERIMENTAL RESULTS

In this section, we compare the performance of the proposed temporal DIRECT mode decision algorithm (denoted as proposed TDMD) with the temporal DIRECT mode decision method proposed in the H.264/AVC encoder (denoted as original TDMD). In the proposed TDMD, the best MV or weight is selected from the co-located block and its eight neighboring blocks for DIRECT mode decision based on the criterion given in (9). The proposed TDMD is only applied to DIRECT 16x16 and B SKIP 16x16. To reduce extra overhead, the DIRECT 8x8 mode uses the original DIRECT mode decision.

QCIF	Foreman
	Claire
	Trevor
	Mobile
	coastguard
CIF	Waterfall
	Stefan
	news
	bus
	Dancer
4CIF	City
	Crew
	Harbour
	Ice
	Soccer

Code Version	JM 12.2
Profile	Main
GOP Structure	IBPBP...
Encoding Frames	199
Frame Rate	30
N_P	2
$N_{B,List0}$	1
$N_{B,List1}$	1
$N_{P,List1}$	2
QP	QPB=QPP+3
RDO	On
Entropy Coding	CAVLC

TABLE II Simulation conditions

We implement these algorithms into the JM encoder JM12.2 to evaluate their performance. The simulation uses fifteen test sequences, covering a wide range of motion activities and various formats (QCIF: 176×144, CIF: 352×288, and 4CIF: 704×576). In the experimental setting, each sequence has 200 frames in simulations for sequence coded with IBPBP structure. The frame rate is 30 frames per second and the quantization parameter for B frames is set as $QP_B=QP_P+3$ [2]. The

experimental setting is summarized in TABLE II.

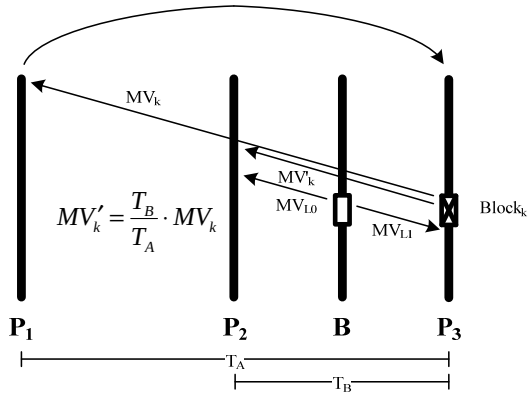

Fig. 3 bilinear interpolation for forward/backward MVs

For simplicity the number of reference frames for motion estimation is $N_P = 2$, i.e., with two reference frame buffers. The numbers of reference frame for B frames are $N_{B,List0} = 1$ and $N_{B,List1} = 1$ respectively, while $N_{P,List0} = 2$ for P frames. Note that the bilinear interpolation is employed to procure the desired MV in the *List 0 reference* frame when the reference number of the *List 1 reference* P frame is 2, as illustrated in Fig. 3.

The performance is compared based upon Bjontegaard Delta PSNR (BDPSNR) and Bjontegaard Delta Bit Rate (BDBR) [4] for QP_P=20, 24, 28 and 32. TABLE III displays the BDPNR and BDBR results, as compared to original TDMD, that shows both cases with and without extra overhead. As demonstrated, the proposed TDMD achieves average 0.22 dB BDPSNR gain and 4.1% of BDBR bit-rate saving when the extra overhead is not considered. When the extra overhead is taken into account, the BDPSNR gain lessens from 0.22 dB to 0.1 dB and the BDBR reduction lessens from 4.1% to 1.6%. The proposed TDMD still outperforms the original TDMD.

As shown in TABLE III, the superiority of the proposed algorithm is evident for fast motion video sequences such as *foreman*, *mobile*, *bus* etc. In these sequences with high motion activities, the occlusion phenomenon occurs often and the MV of the co-located block cannot usually represent the true motion of the current encoding block, leading serious prediction errors. To obtain further insight, Fig. 4 compares RD performance for various QPs, carried on *mobile* to show its superiority over original TDMD algorithm.

No matter how, the advantage of the proposed TDMD becomes lost for video sequences with slow motion activities like *claire*. This is because that most areas in *claire* sequence are homogeneous and stationary, and the bidirectional prediction using MV of the co-located block for DIRECT mode usually gives very good coding efficiency, compared to its neighboring blocks. As a result, extra overhead for MV and weight selection degrades its coding performance severely.

QPP20,24,28,32		BDPSNR (dB)		BDBR (%)	
		No Overhead	Overhead	No Overhead	Overhead
QCIF	Foreman	0.396	0.265	-8.480	-5.671
	Claire	0.268	-0.115	-4.965	2.284
	Trevor	0.260	0.097	-4.290	-1.591
	Mobile	0.365	0.313	-4.428	-3.783
	coastguard	0.203	0.121	-3.583	-2.143
CIF	Waterfall	0.462	0.329	-10.859	-7.734
	Stefan	0.192	0.134	-2.751	-1.901
	news	0.244	-0.010	-4.392	0.207
	bus	0.347	0.271	-5.211	-4.052
	Dancer	0.101	-0.039	-2.059	0.779
4CIF	City	0.109	0.049	-2.888	-1.276
	Crew	0.049	-0.015	-1.420	0.390
	Harbour	0.061	0.008	-1.141	-0.142
	Ice	0.141	-0.018	-3.955	0.494
	Soccer	0.028	-0.023	-0.652	0.477
Average		0.215	0.091	-4.072	-1.577

TABLE III BDPSNR and BDBR comparison

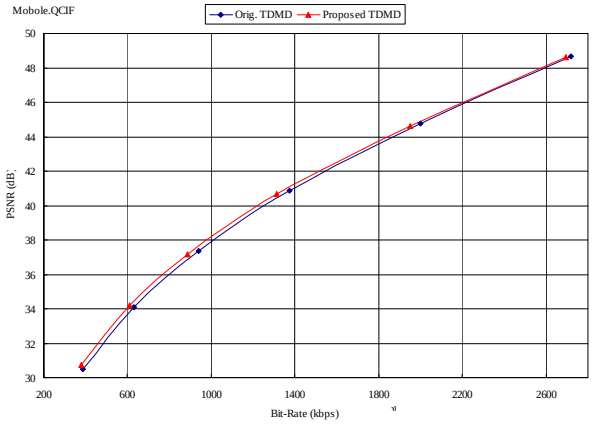

Fig. 4 Rate distortion curve on *mobile*

V. CONCLUSION

In this paper, we suggest a temporal DIRECT mode decision algorithm for H.264/AVC inter B frame video coding to enhance the coding performance. The proposed algorithm uses MVs of the co-located block as well as its eight neighboring blocks for DIRECT mode decision. In addition, the weight selection is also considered for occlusion areas. The experimental results reveal that average PSNR gain of 0.1 dB, or corresponding to average 1.6% of bit-rate reduction can be achieved, compared to the temporal DIRECT mode decision proposed in H.264/AVC.

REFERENCES

[1] A. Vectro, C. Christopoulos, and H. Sun, “Video transcoding architectures and techniques: An overview,” *IEEE Signal Processing Magazine*, vol. 20, pp. 18-29, March 2003.
[2] M. Flierl and B. Girod, “Generalized B pictures and the draft H.264/AVC video compression-standard,” *IEEE Trans. Circuits Syst. Video Technol*., vol. 13, no. 7, pp. 587-597, July 2003.
[3] A. M. Tourapis, F. Wu and S. Li, “Direct mode coding for bipredictive slices in the H.264 standard,” *IEEE Trans. Circuits Syst. Video Technol*., vol. 15, no. 1, pp. 119-126, January 2005.
[4] G. Bjontegaard, “Calculation of average PSNR difference between RD curves”, ITU-T Q.6/16, Doc. VCEG-M33, April 2001.