ENHANCED DIRECT MODE CODING FOR BI-PREDICTIVE PICTURES

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

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, ylu, dbzhao, wgao, swma}@jdl.ac.cn

ABSTRACT

The direct mode coding in the bi-predictive pictures (B-pictures) can efficiently improve the performance of bi-predictive coding, because it exploits the temporal correlation by bi-directional prediction from both forward and backward reference pictures, and meanwhile it does not require any bits for coding the motion vectors. Accordingly, how to accurately calculate the motion vectors in terms of direct mode coding is very important to obtain better prediction values. In other words, it is more desirable to obtain the true motion vectors, because the derived motion vectors need not be coded. The traditional direct mode coding usually derives the motion vector of the current block by scaling the motion vector of the co-located block in the subsequent reference picture. It is based on the assumption that the near blocks would have the same motion trajectory. In this paper, we present an improved method based on combining motion vector tracking technique and spatial motion vector prediction technique. The proposed method can accurately calculate the motion vectors for the direct mode coding.

1. INTRODUCTION

In terms of the video coding, the fundamental problem is about how to utilize the motion-compensated prediction. It has been proven that the bi-directional prediction used for bi-predictive coded pictures (B-pictures) is a very efficient tool to enhance coding efficiency. In H.263 and MEPG-4 standard [1][2], the direct mode has been utilized in B-pictures, which does not require any bits for coding the motion vector data in that the motion vector of the current block can be derived directly from that of its co-located block in the temporally subsequent reference picture. Sometimes, the derived forward and backward motion vectors are so accurate that they need not be corrected with any differential motion vectors. In this case, the direct mode can be efficiently utilized so as to save the bits for coding the motion vectors. In general, direct mode coding can efficiently improve the coding performance of B-pictures.

The direct mode that derives the bi-directional motion vectors from its co-located block is referred to as the Temporal Direct Mode (TDM) in the H.264/AVC standard [3]. The major difference of direct mode between H.264/AVC and the other standards such as H.263 and MEPG-4 is that H.264/AVC can use multiple reference pictures in prediction. Accordingly, besides the traditional TDM used as forward/backward, it also supports the forward/forward and backward/backward [4]. Considering the generosity, the discussion in this paper is only focused on the forward/backward case, whereas the described methods can be easily extended for the other cases. Fig. 1 illustrates how to derive the forward and backward TDM motion vectors, in which MV_F points to reference pictures List0 (forward) and MV_B points to reference picture List1 (backward) in H.264/AVC 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 picture and the temporally previous reference picture, td denotes the temporal distance between the temporally previous reference picture and temporally subsequent reference picture, and MV_D denotes the motion vector of the co-located block in the temporally subsequent reference picture.

TDM can efficiently exploit the temporal redundancy between adjacent pictures. However, the motion vector of the current block derived from that of its co-located block in the temporally subsequent reference picture is not accurate enough, because the current block is not always located in the motion trajectory of its co-located block, particularly when the current block and its co-located block in temporally subsequent reference picture belong to different objects with different motions. Moreover, some other coding conditions, such as more B pictures utilized, video sequence with small and/or fast motion, and half pixel accuracy interpolation, etc, also affect the accuracy of the derived motion vector for TDM coding.

A further problem of traditional direct mode coding lies in that the motion vector of some co-located blocks in the temporally subsequent reference picture cannot be obtained in some cases, such as uncovered boundaries, overlapped blocks and the luminance change in adjacent pictures, etc, which will also lead to many blocks coded with intra mode. If we fail in the derivation of motion vector for TDM and only set it zero, the accuracy of the derived TDM motion vector will be deteriorated.

To tackle the above problems, in this paper, we propose a new method based on the combination of the projection and spatial prediction techniques so as to accurately derive the motion vector for direct mode coding. In the following section, we are focused on presenting the proposed method about how to derive the motion vector for direct mode coding. Then the simulated results are presented.

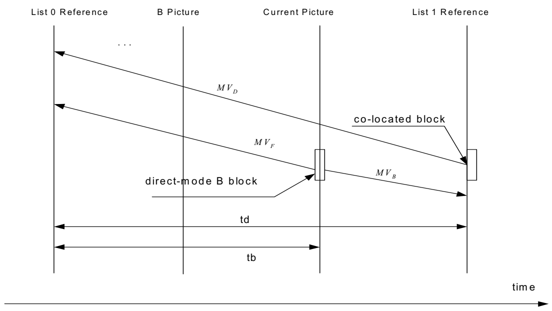


Fig.1. Temporal direct mode used in H.264/AVC

2. PROPOSED DERICT MODE CODING

2.1 Motion vector tracking

Motion vector tracking technique was introduced to improve the interframe coding of motion vectors in [5], and reduce the match searching numbers for motion estimation because it can track more accurate candidate motion vector as a starting point in the motion estimation algorithm [6]. As shown in Fig. 2, motion vector tracking technique can track the true motion vector V_t of the block M_t from one picture P(t) to another picture P(t+1). Compared with deriving the motion vector from the co-located block in previously coded picture, this kind of temporal projection can better track the true motion trajectory of the object between temporally adjacent pictures.

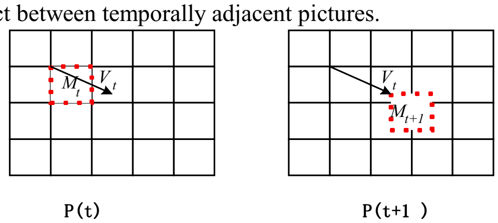


Fig 2. Motion vector tracking from the picture P(t) to the temporally next picture P(t+1)

Regarding the B picture, the motion vector of the block in temporally subsequent picture can also track the motion vector from temporally subsequent picture to temporally previous picture. That is to say, we can improve the accuracy of the motion vector derivation for TDM by scaling the motion vector pointing to one block Mt-1 in the temporally previous reference picture, of the block Mt+1 in the temporally subsequent picture when it tracks through the current block Mc in B picture, as shown in Fig. 3.

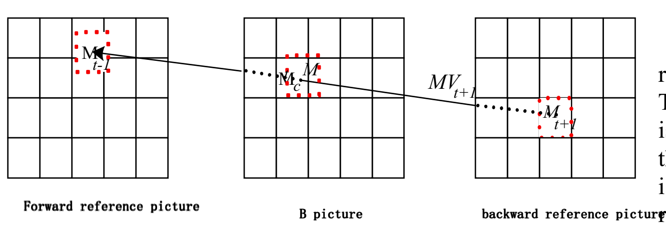


Fig. 3. Motion vector derivation of the block Mc in B-picture

To further improve the accuracy of the derived motion vectors for TDM, and to make the current block Mc cover at least 1/4 corner of the actual moving block Mt in the motion trajectory, we assume that the motion vector of the block Mt+1 starts from the center of the block, as shown in Fig. 3. Then, we can generate the motion vector trajectory in terms of each block in the temporally subsequent reference picture that starts from its center. Accordingly, we can derive the motion vector of the block for TDM as follows:

$$CP(x,y) = M_{t+1}(x-n/2, y-n/2) + ((tb-td)/td) \times MV_{t+1}, \quad (3)$$

$$M_C = CP(x,y)/n \quad (4)$$

$$MVB_{Mc} = ((tb-td)/td) \times MV_{t+1}, \quad (5)$$

$$MVF_{Mc} = (\text{tb/td}) \times MV_{t+1}, \quad (6)$$

where $n \times n$ is the block size , MV_{t+1} is the motion vector of the block M_{t+1} in the temporally subsequent reference picture that points to the block M_{t-1}, $M_{t+1}(x, y)$ and $M_{t+1}(x-n/2, y-n/2)$ denotes the up-left corner and center of the block M_{t+1}, respectively, $CP(x, y)$ is the pixel position of the point of intersection between motion vector MV_{t+1} and the current B picture, namely the center of M_t located on the trajectory of motion vector MV_{t+1}, M_c is the block in B picture which covers at least 1/4 corner of the actual moving block M_t in the motion trajectory MV_{t+1}, MVF_{Mc} and MVB_{Mc} denote the forward and backward motion vectors of M_c, respectively, / is integer division with truncation of the result toward zero.

2.2 Spatial prediction technique

In H.264/AVC, another type of direct mode, i.e. Spatial Direct Mode (SDM) [7], is also utilized, wherein the motion vectors are derived from the motion vectors of spatially adjacent blocks. The spatial 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. 4. SDM can efficiently exploit the spatial correlation and sometimes the temporal redundancy by considering whether a block is stationary or not according to the motion vector of the co-located block. This method can achieve good performance for smooth and/or scene-changed sequences, which requires less memory than the temporal direct mode.

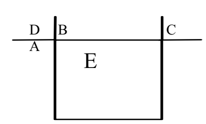


Fig.4. Median prediction of motion vectors

In H.264/AVC, when the co-located block in the subsequent reference picture is coded with intra mode, the corresponding TDM motion vectors has to be set as zero. However, this method is inefficient in most cases. To tackle this problem, we introduce the spatial prediction scheme in SDM to the field of TDM. That is to say, similar to the SDM that exploits the temporal redundancy of the motion vector of the co-located block by considering whether a block is stationary or not, TDM also can consider partially exploiting the spatial redundancy, especially for the scene-changed cases.

This method is similar to the skip mode in forward predictive coded pictures (P-pictures), whose motion vector is obtained from median predictor and only used based on the 16×16 macroblock. However, the minimum block size of luma motion compensation is as small as 4×4 in H.264/AVC, and 8×8 in MPEG-4 and H.263. Accordingly, we can further exploit the correlation in the macroblock with different motion represented by motion vectors of the blocks less than 16×16 in size in temporally subsequent picture in the proposed scheme.

2.3 Combination of motion tracking and spatial prediction

As for the motion vector tracking technique, not every block in B picture owns one motion vector tracking through it, because some other blocks may occupy more than one derived motion vectors and also some blocks contain no motion vector derived from the temporally subsequent reference picture when it is coded in intra mode. Accordingly, there exist three types of blocks: without derivation motion vector, only one derivation motion vector and more than one derivation motion vector. Based on the proposed combination of motion tracking and spatial prediction, we can deal with these cases as follows:

1) If only one derivation motion vector MV_{t+1} tracks through the current block M_c in the B picture. We can derive its forward motion vector MVF_{Mc} and backward motion vector MVB_{Mc} for TDM according to equation (5) and (6).
2) If more than one derivation motion vector $MV_{t+1}(i)$ tracks through the current block M_c in the B picture, we will select the motion vector of the block M_t in motion trajectory which covers the largest part of the current block M_c as the derivation motion vector MV_{t+1} according to (7), and then we can derive MVF_{Mc} and MVB_{Mc} for TDM according to equation (5) and (6).

$$MV_{t+1} = \mathbf{min}((CP_i(x) - M_c(x-n/2))^2 + (CP_i(y) - M_c(y-n/2))^2) \quad (7)$$

Where $CP_i(x)$, $CP_i(y)$ is the horizontal and vertical component of the point $CP_i(x,y)$ of intersections between motion vector $MV_{t+1}(i)$ and the current B picture.

3) If no derivation motion vector track through the current block M_c in B picture because some other blocks occupy more than one derivation motion vector and no motion vector can be obtained for some blocks in the temporally subsequent reference picture when they are coded in intra mode. In this case, we can derive TDM motion vectors by spatial motion vector prediction. The forward and backward motion vector MVF_{Mc}, MVB_{Mc} for TDM are separately obtained from the coded forward motion vectors of its neighboring blocks in the current B picture by median prediction technique. However, considering the scene-changed case, if MVF_{Mc} or MVB_{Mc} is not available because all forward or backward motion vectors of its neighboring blocks are not available, the single prediction or bi-prediction can be used for direct mode. if both MVF_{Mc} and MVB_{Mc} are not available because all the forward and backward motion vectors of the neighboring blocks are not available, we set MVF_{Mc} and MVB_{Mc} zero.

3. EXPERIMENTAL RESULTS

The proposed method is implemented based on the H.264/AVC reference software JM6.1e [8]. The test sequences in the CIF format include *foreman, coastguard, paris, tempete, mobile,* and *flower* with 30fps. The test conditions are showed in Table 1. To evaluate the average PSNR vs. bit-rate, we employ the method described in [9], which is widely used during H.264/AVC development. The detailed results are show in Table 2. In the testing, only one forward reference picture and one backward reference picture are used. Thus, the coding results can reveal the performance of the conventional forward/backward case of B-picture coding. In other words, it demonstrates the proposed algorithm is not only suitable for H.264/AVC, but also suitable for H.263 and MPEG-4.

Table 2 shows that the performance of our method is generally better than TDM in H.264/AVC. The maximum PSNR gain is up to 2.056dB in terms of B-pictures and 0.389dB for total pictures when IBPBP… GOP structure is used. The maximum PSNR gain is up to 0.995dB in terms of B-pictures and 0.371dB for total pictures when IBBPBBP…GOP structure is used.

Fig. 5 illustrates that the number of blocks coded with direct mode in each B frame is always larger than that of the TDM technique in H.264/AVC when IBPBP… GOP structure is used. In other words, it verifies that the improving of the accuracy of the derived motion vector for TDM can efficiently increase the numbers of the direct mode in each frame, and consequently, obtain the better prediction values.

Fig. 6 shows the rate-distortion curves of B pictures for the test sequences of mobile and tempete when IBPBP… GOP structure is used. At high and middle bit rates, the proposed method can significantly improve the coding efficiency, whereas at the low bit rate, the performances between the proposed method and TDM almost are very close, in that the effectiveness of direct mode decreases when motion vector in the temporally subsequent reference picture is not accurate enough.

Table 1. Test condition

MV resolution	1/4 pel
Hadamard	ON
RD optimization	ON
Search Range	±16
Restrict Search Range	2 (no restrictions)
Reference Frames	1
Symbol Mode	(CAVLC)
GOP structure	IBPBP… and IBBPBBP…
QP	20,28,34,40

4. CONCLUSIONS

For small and/or fast motion video sequences, as we known, the TDM motion vector of the current block cannot be accurately derived from the motion vector of its co-located block, because it cannot track the true motion vector of moving object. The proposed method can further exploit the temporal correlation, in that it introduces the more accurate derivation of motion vector rather than TDM. Furthermore, the proposed method also integrates a technique that can exploit the spatial redundancy by spatial motion vector prediction.

5. ACKNOWLEDGEMENT

This work has been supported in part by National Science Foundation of China (69789301), by National Hi-Tech Development Program of China (2001AA114160), and by National 973 Program of China (2001CCA03300).

6. REFERENCES

[1] ITU-T Recommendation H.263, "Video coding for low bit-rate communication," Nov. 1995

[2] ISO/IEC JTC1/SC29/WG11 N2502, FDIS of ISO/IEC 14496-2, "Generic coding of audio-visual objects: Part 2- Visual," (MPEG-4), Nov. 1998.

[3] 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-.les.org/jvt-experts/2003 03 Pattaya/ JVT-G050r1.zip.

[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] J. Yeh, M. Vetterli, M. Khansari ,"Motion Compensation of Motion Vectors", International Conference on Image Processing Vol. 1, pp. 574-578, 1995

[6] Ismaeil R. Ismaeil, Alen Docef, Faouzi Kossentini, Rabab K. Ward: Efficient Motion Estimation Using Spatial and Temporal Motion Vector Prediction. ICIP (1) 1999: 70-74.

[7] 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.

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

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

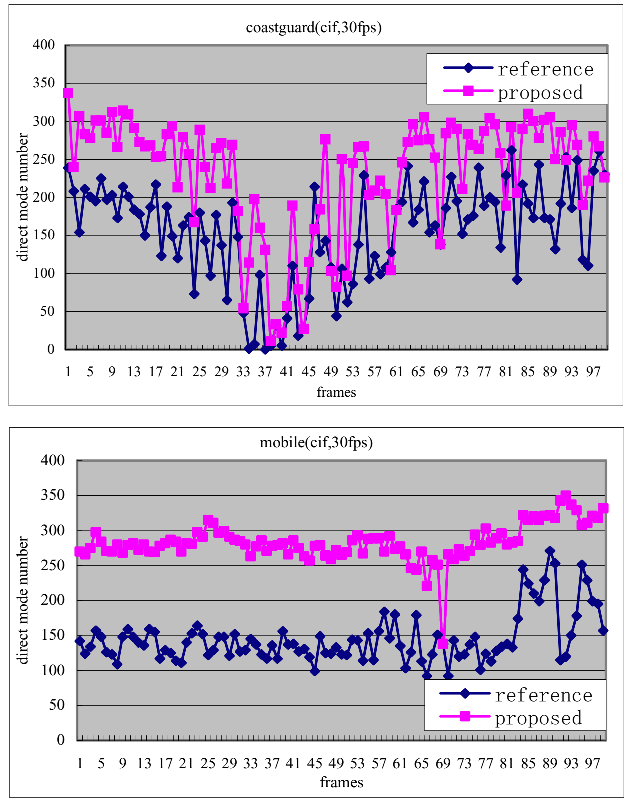
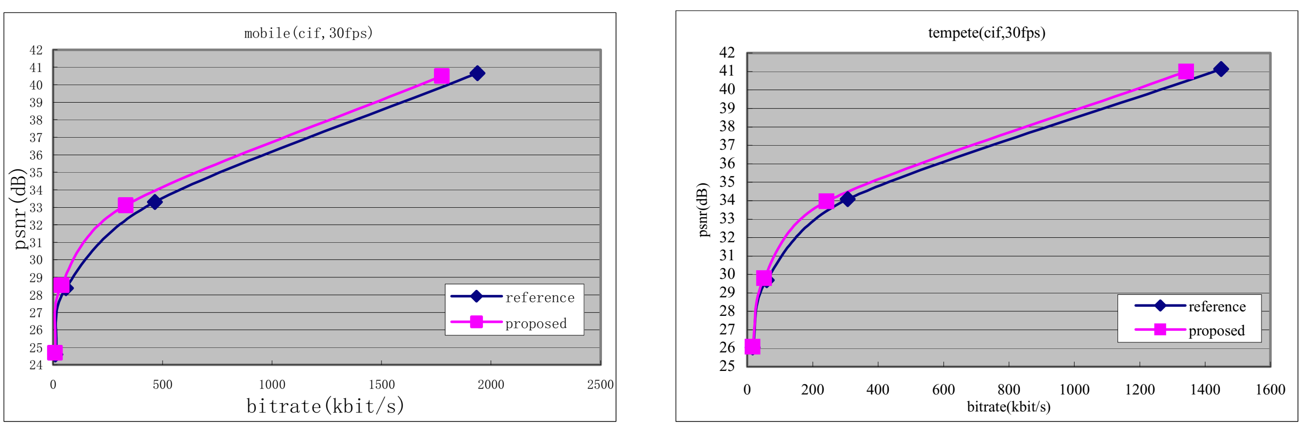


Fig.5. Number of blocks coded with direct mode in each frame.



(a) (b)

Fig. 6. Rate-distortion curves for sequences (a) mobile and (b) tempete

Table 2. Performance comparisons between proposed and TDM.

Video sequences		foreman	coastguard	Mobile	Paris	tempete	flower
all pictures with IBPBP…	*Ave PSNR gain*	*0.057*	*0.145*	*0.389*	*0.165*	*0. 235*	*0. 269*
	Ave BR Saving	*1.03%*	*2.35%*	*5.45%*	*2.53%*	*3.61%*	*3.43%*
B pictures with IBPBP…	*Ave PSNR gain*	*0.338*	*0.983*	*2.056*	*0.731*	*0.923*	*1.730*
	Ave BR Saving	*5.75%*	*14.45%*	*21.76%*	*10.92%*	*13.71%*	*17.53%*
all pictures with IBBPBBP…	*Ave PSNR gain*	*0.053*	*0.129*	*0.371*	*0.114*	*0. 205*	*0. 204*
	Ave BR Saving	*0.89%*	*1.93%*	*5.32%*	*1.81%*	*3.29%*	*2.55%*
B pictures with IBBPBBP…	*Ave PSNR gain*	*0.142*	*0.435*	*0.995*	*0.308*	*0.504*	*0.659*
	Ave BR Saving	*2.51%*	*5.63%*	*12.88%*	*4.68%*	*7.71%*	*7.45%*