Tagged Multi-hypothesis Motion Compensation Scheme for Video Coding

Lei Chen[*#1], Ronggang Wang[*2], Siwei Ma[#3]

[*]*Peking University, Shenzhen Graduate School, Shenzhen, China*

[#]*Institute of Digital Media, Peking University, Beijing, China*

[1]*leic17@sz.pku.edu.cn*

[2]*rgwang@pkusz.edu.cn*

[3]swma@pku.edu.cn

***Abstract*— Accuracy of prediction block (PB) plays a very important role in improving the coding performance. In this paper, we propose tagged multi-hypothesis motion compensation scheme (TMHMC) for inter frames to improve the accuracy of PB. TMHMC not only makes use of temporal correlation between frames but also the spatial correlation as motion vectors of adjacent blocks are used to derive the PB. For entropy coding process, only one motion vector and a tag indicating which adjacent block is used are coded in bit-stream. Adding TMHMC scheme as an additional mode in MPEG internet video coding (IVC) platform, the bitrate saving is up to 12% at the same objective quality compared with anchor. Average bitrate saving is close to 6% over all test sequences. In addition, we also implement the conventional multi-hypothesis motion compensation (MHMC) scheme. 3% bitrate is further saved on average by TMHMC compared with conventional MHMC.**

***Index Terms*—Multi-hypothesis, motion compensation, video coding, MPEG IVC, tag**

I. Introduction

In video coding technology, spatial relation characteristics in a frame and temporal relation characteristics between frames are exploited to reduce the amount of actual encoded data. Accuracy of PB is important to improve the coding performance. For inter frames, motion compensation process, which makes use of temporal relation characteristics, is conducted to get the best PB and minimize the actual encoded data.

In conventional motion compensation, PB of current block is derived through motion vector searched in motion estimation [1-3]. For inter blocks in P frames, only one motion vector is searched and thus, only one PB is obtained which results in limited accuracy of PB. For bi-directional blocks in B frames, two motion vectors are searched and two temporal PBs are derived by these two motion vectors. In this way, the accuracy of the final PB is improved. Generalized B frame [4] uses the same idea to improve the prediction accuracy of inter blocks in P frames by MHMC [5-6]. However, for bi-directional B frames, low delay B frames and generalized B frames, two motion vectors are required to be searched and coded in bit-stream which increases the number of bits consumed.

In this paper, we propose TMHMC scheme to improve the accuracy of PB with less bitrate cost. Similar to generalized B frames, two temporal PBs are derived and the final PB is the average of these two temporal PBs. However, different with generalized B frames, we only code one motion vector and one tag instead of two motion vectors to reduce the bitrate cost. In our proposed scheme, not only the temporal correlation characteristics between frames are used, but also the spatial correlation characteristics in a frame are exploited. Motion vectors of adjacent blocks are used to get the first temporal PB. The other temporal PB is derived by motion vector searched through joint motion estimation depending on the first temporal PB. For entropy coding process, the motion vector searched in joint motion estimation and the tag indicating the adjacent block used to derive the first temporal PB are coded in bit-stream.

MPEG internet video coding (IVC) platform ITM4.0 [7] is taken as a benchmark. TMHMC scheme is implemented in it as a new mode for inter frames. Rate-distortion optimization (RDO) process utilized in H.264 [8] and HEVC [9] coding standard is used to select the best mode. The bit-rate reduction of modified ITM4.0 implemented with TMHMC scheme is up to 12% at the same objective quality. Average reduction is close to 6% over all test sequences. What’s more, compared with conventional MHMC scheme, 3% bitrate is further saved by TMHMC.

The rest of this paper is organized as follows. Section 2 presents the inter-frame coding framework implemented with our TMHMC scheme. Section 3 describes TMHMC scheme in detail. Experimental results are presented in section 4. Finally we conclude this paper in section 5.

II. PROPOSED INTER-FRAME CODING FRAMEWORK

The proposed scheme is integrated in MPEG IVC platform as an inter-frame coding mode. Fig. 1 depicts the proposed framework. To obtain the final PB, two inter prediction paths are conducted. The first one is the conventional motion compensation process which exists already in IVC reference software and the other one is our proposed TMHMC scheme. Each path drives a corresponding PB and the one that makes the minimum RD-cost value is selected. After motion compensation process, the residual block is obtained through subtracting the PB from the original one. Then a 2D transform is performed on the residual block. Finally, quantization and entropy coding is performed as usual. In entropy coding process, only one motion vector and one tag are coded in bit-stream when the TMHMC mode is selected.

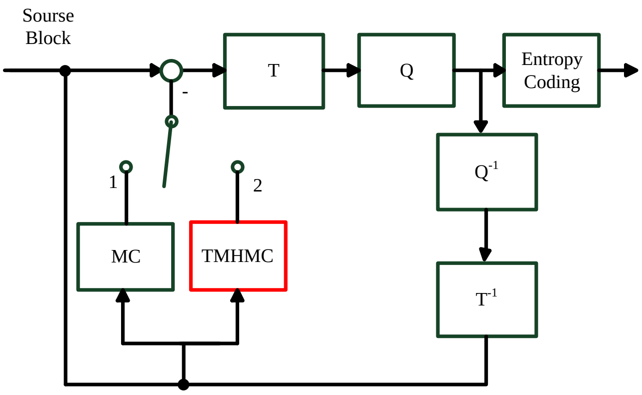

Fig. 1 Proposed inter-frame coding framework.

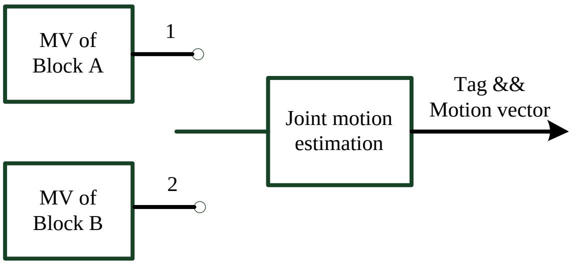

Fig. 2 Detail framework of TMHMC.

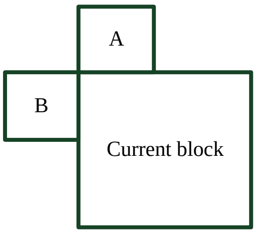

Fig. 3. Adjacent blocks of current block.

The detail framework of TMHMC scheme is further depicted in Fig. 2. Block A represents the above adjacent block of the current block and their positional relationship is presented in Fig. 3. Block B is the left adjacent one. Both motion vectors of Block A and Block B are utilized to do joint motion estimation process and obtain the final PB in TMHMC mode. In TMHMC, two temporal PBs should be derived to obtain the final PB. The first one is indicated by motion vector of Block A or Block B. The final motion vector indicating the first temporal PB is selected through RDO process. The switch button in Fig. 2 stands for the RDO process. The other temporal PB is derived through joint motion estimation which refers to the first temporal PB. Assuming that the first temporal PB derived by motion vector of Block A is called $PB1_A$ and the first temporal PB derived by motion vector of Block B is called $PB1_B$, $PB1_A$ and $PB1_B$ are used separately in joint motion estimation process to derive the second temporal PB, $PB2_A$ and $PB2_B$. In this way, two pairs of temporal PBs ($PB1_A$, $PB2_A$) and ($PB1_B$, $PB2_B$) are obtained and can be used to derive the final PB. Through RDO process, the pair of temporal PBs that makes the minimum RD-cost value is selected to derive the final PB. After RDO process, the tag indicating which adjacent block (Block A or Block B) is used and the motion vector obtained through joint motion estimation process are determined and coded in bit-stream. Section 3 presents the detailed TMHMC scheme.

III. TAGGED MULTI-HYPOTHESIS MOTION COMPENSATION SCHEME

As mentioned before, the final PB in TMHMC scheme is derived by two temporal PBs, which is similar to bi-directional mode in B frame. We define the two temporal PBs as P1 and P2 and they can be derived by two motion vectors MV1 and MV2. MV1 is selected through motion vectors of adjacent blocks and MV2 is searched through joint motion estimation process. The final PB, named P, is the average of P1 and P2. Since both motion vectors of above adjacent block and left adjacent block are utilized, the joint motion estimation process should be conducted twice to find which one is better. For entropy coding, arithmetic coding method is utilized to code the motion vector difference of MV2 and the tag of MV1. The following subsections describe how to derive MV1, MV2 and the final PB.

A. Derivation of MV1

Motion vector of above adjacent block (block A) or left adjacent block (block B) shown in Fig. 3 is directly used as the first motion vector MV1. For each adjacent block, a corresponding temporal PB ($PB1_A$ or $PB1_B$) is obtained and used in derivation process of the second motion vector. The one that makes better final PB is selected as MV1.

B. Derivation of MV2

MV2 is derived through joint motion estimation process which refers to the first temporal block ($PB1_A$ or $PB1_B$). The joint motion estimation process is similar to SYM mode [10] of B frames in AVS coding standard [11] and bi-directional mode in H.264 and HEVC coding standard. Without loss of generality, the case of one reference frame is used in this discussion and the next subsection.

The joint motion estimation process to derive MV2 for a coded block S is performed by minimizing the Lagrangian cost function:

$$J(MV1, MV2) = D_{SAD}(MV1, MV2) + \lambda_{SAD} \times R(MV2 - MV2_{pred}) \quad (1)$$

where, *MV1* denotes the motion vector of adjacent block. *MV2* represents the target motion vector and is calculated by minimizing Eq.(1). $R(MV2\text{-}MV2_{pred})$ denotes the number of bits to code the difference between MV2 and its predictor. The distortion term is further defined as Eq.(2).

$$D_{SAD}(MV1, MV2) = \sum_{(x,y)} |S(x,y) - (S_{ref}(x+MV2_x, y+MV2_y) + S_{ref}(x+MV1_x, y+MV1_y)) >> 1| \quad (2)$$

where, (x, y) is the coordinates of pixels in the current block. S_{ref} represents the reference picture. $MV1_x$ and $MV1_y$ are the x and y component of MV1, and $MV2_x$ and $MV2_y$ are the x and y component of MV2.

Derivation process of predictor of MV2 is the same as the motion vector prediction method proposed in [12]. In entropy coding, the difference between MV2 and its predictor is coded by arithmetic coding method in MPEG IVC.

C. Derivation of the final PB

Since motion vectors of block A and block B are candidates of MV1, the joint motion estimation process should be conducted twice. For the first time, the motion vector of the above adjacent block is used as MV1. And for the second time, the motion vector of the left adjacent block is used as MV1. RDO is utilized to select the better one as the final MV1. After MV1 and MV2 are determined, the first and second temporal PB, P1 and P2, can be determined in the corresponding reference picture (shown in Fig. 4). The final PB, named P, is the average of P1 and P2 and is calculated through Eq.(3).

$$\begin{aligned} P(x,y) = (&P1(x + MV1_x, y + MV1_y) \\ &+ P2(x + MV2_x, y + MV2_y)) >> 1 \end{aligned} \quad (3)$$

where, *(x,y)* is the coordinates of pixels in current block. $MV1_x$ and $MV1_y$ are the x and y component of MV1. $MV2_x$ and $MV2_y$ are the x and y component of MV2. *P1* and *P2* are the first and second temporal PBs.

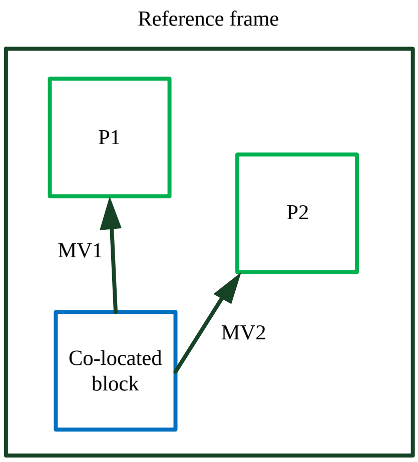

Fig. 4 Multi-hypothesis predictors of current block.

D. Entropy coding

For entropy coding, since motion vectors of adjacent blocks are already derived in both encoder and decoder, only the tag indicating the motion vector used as MV1 is coded. Besides, the difference between MV2 and its predictor is also coded. The arithmetic coding algorithm is used in entropy coding.

IV. Experimental Results

For all the tests, MPEG IVC ITM4.0 is used as anchor and test platform. All the test sequences in MPEG IVC call for proposals (CfP) are tested and the resolution ranges from 4K to WQVGA. The common test condition in low-delay case (Constraint Set 2) of MPEG IVC is used for all the tests.

The proposed TMHMC scheme is integrated into the inter-frame coding framework of MPEG IVC software ITM4.0 as an additional mode, based on the framework shown in Fig. 1 and Fig. 2. To further compare the performance of TMHMC with conventional MHMC, conventional MHMC scheme is also implemented in ITM4.0 as an additional mode. In conventional MHMC, two motion vectors indicating the two temporal PBs is required to be searched and coded in bit-stream.

TABLE I

BD-Rate Performance of MHMC and TMHMC

Resolution	**Test sequences**	**MHMC vs. anchor**	**TMHMC vs. anchor**
Class A 4K	Traffic	-1.3%	-2.7%
	PeopleOnStreet	-1.6%	-5.9%
Class B 1080P	Kimono	2.3%	-3.2%
	ParkScene	-1.1%	-3.4%
Class C WVGA	BasketballDrill	-1.7%	-4.3%
	BQMall	-3.2%	-7.0%
	PartyScene	-2.2%	-4.4%
	RaceHorses	-5.5%	-9.9%
Class D WQVGA	BasketballPass	-2.7%	-6.8%
	BQSquare	-8.7%	-11.9%
	BlowingBubbles	-2.7%	-4.7%
	RaceHorses	-4.2%	-8.5%
Class E 720P	FourPeople	-0.9%	-4.5%
	Johnny	-3.7%	-6.2%
	KristenAndSara	-1.1%	-4.0%
4K		-1.4%	-4.3%
1080P		0.6%	-3.3%
WVGA		-3.2%	-6.4%
WQVGA		-4.6%	-8.0%
720P		-1.9%	-4.9%
Average		-2.6%	-5.8%

The test results are listed in Table 1. The first and second columns respectively show the resolution and name of each test sequences. The third column and the forth column present the coding performance of conventional MHMC scheme and our proposed TMHMC scheme compared with anchor. As shown in Table 1, TMHMC scheme is efficient and saves BD-rate for all the test sequences. The BD-rate reduction is up to 12% for BQSquare sequence in class D. Besides, average 5.8% BD-rate is achieved by our TMHMC scheme. Compared with TMHMC scheme, MHMC is not so efficient since two motion vectors need to be coded in bit-stream. As Table 1 indicates, MHMC scheme saves BD-rate for most of the test sequences except for Kimono in class B. The BD-rate loses 2.3%. TMHMC scheme improves the performance of MHMC scheme and 3.2% BD-rate is further saved on average. In Fig. 5, the RD curves on the sequences PeopleOnstreet in class A, BQMall in class C, RaceHorses in class D and FourPeople in class E are provided.

[13] and our TMHMC are both about MHMC, but they focus on different application scenarios. The former is a low-cost MHMC method, whereas our method is to improve the coding efficiency.

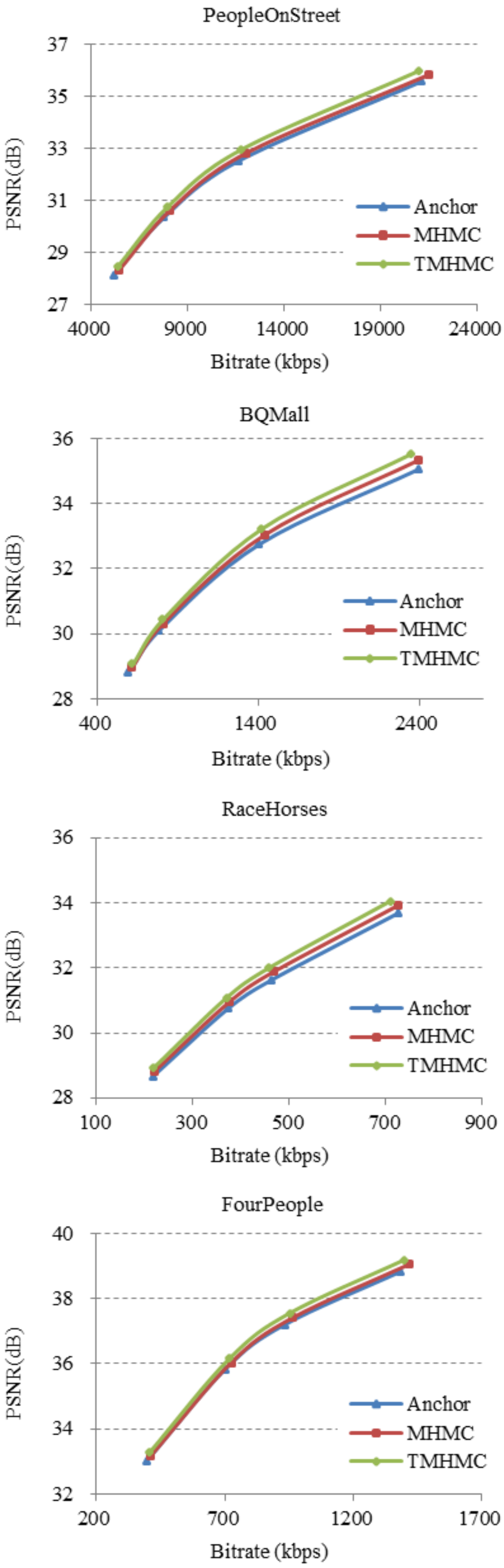

Fig. 5. RD curve performance.

Finally, for the complexity aspect of the proposed TMHMC, the encoding time is increased since joint motion estimation is required to be conducted twice. However, the complexity of TMHMC is comparable to that of MHMC. In the near future, we plan to reduce the complexity by using fast mode decision and fast motion estimation methods.

V. CONCLUSION

In this paper, a new motion compensation scheme TMHMC is proposed. Compared to conventional motion compensation, it not only makes use of temporal correlation characteristics but also the spatial correlation characteristics. Similar to bi-directional mode in conventional B frames, the PB in TMHMC is derived by two motion vectors. However, only one tag and one motion vector need to be coded in bit-stream instead of two motion vectors compared with bi-directional mode in B frames. So, the accuracy of PB maintains with less bitrate cost. Experimental results show that the proposed scheme significantly improves the coding performance of conventional motion compensation and conventional MHMC.

ACKNOWLEDGMENT

This work was supported in part by the National High-tech R&D Program of China (863 Program, 2012AA010805)，National Science Foundation of China(61322106, 61103088，61390515)

REFERENCES

[1] S. Zhu and K.-K. Ma, "A new diamond search algorithm for fast block-matching motion estimation", *IEEE Trans. Image Processing*, vol. 9, pp.287 -290 2000.

[2] L. M. Po and W. C. Ma, "A novel four-step search algorithm for fast block motion estimation", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp.313 -317 1996.

[3] Z. Chen, J. Xu, Y. He and J. Zheng, "Fast integer-pel and fractional-pel motion estimation for H.264/AVC", *Journal of Visual Communication and Image Representation*, vol. 17, pp.264 -290 2006.

[4] M. Flierl and B. Girod, "Generalized B pictures and the draft JVT/H.264 video compression standard", *IEEE Trans. Circuits Syst. Video Technol*., vol. 13, pp.587 -597 2003.

[5] G. J. Sullivan, "Multi-hypothesis motion compensation for low bit-rate video coding", Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), pp.437 -440 1993.

[6] M. Flierl, T. Wiegand and B. Girod, "Rate-constrained multi-hypothesis motion-compensated prediction for video coding", Proc. IEEE Int. Conf. Image Processing, vol. 3, pp.150 -153 2000.

[7] Ronggang Wang, Xianguo Zhang, Siwei Ma, Jianwen Chen, Cliff Reader, "Internet Video Coding Test Model (ITM) Version 4.0", ISO/IEC JTC1/SC29/WG11 N13353 January 2013, Geneva, Switzerland.

[8] T. Wiegand, G. J. Sullivan, G. Bjntegaard and A. Luthra, "Overview of the H.264/AVC video coding standard", *IEEE Trans. Circuits Syst. Video Technol*., vol. 13, no. 7, pp.560 -576 2003.

[9] G. J. Sullivan, J. -R. Ohm, W. -J. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard", *IEEE Trans. Circuits Syst. Video Technol*., vol. 22, pp.1648 -1667 2012.

[10] X. Ji, D. Zhao, F. Wu, Y. Lu and W. Gao, "B-picture coding in AVS video compression standard," *Signal Process.: Image Communication*, vol.23, Issue 1, pp.31-41 2008.

[11] L. Fan, S. Ma and F. Wu, "An overview of AVS video standard", Proc. IEEE Int. Conf. Multimedia and Expo (ICME), pp.423 -426 2004.

[12] Hao Lv, Ronggrang Wang, Lei chen “A direction and distance based motion vector prediction method for internet video coding”, ISO/IEC JTC1/SC29/WG11 M27965 January 2013, Geneva, Switzerland.

[13] Lei Chen, Ronggang Wang, Siwei Ma, "Low-cost multi-hypothesis motion compensation for video coding," SPIE Visual Information Processing and Communication V, 2014.