New Merge Mode Decision in High Efficiency Video Coding (HEVC)

Chang-Yi Hong, Hui-Jane Hsieh* and Yinyi Lin
Department of Communication Engineering
National Central University, Taiwan
*Department of Information Management
Chien Hsin University of Science and Technology, Taiwan
yilin@ce.ncu.edu.tw

***Abstract*—The advanced motion vector prediction (AMVP) used in the newly developed High Efficiency Video Coding (HEVC) employs five MV candidates from both spatial and temporal domains for merge mode decision. The merge index is then entropy coded using a simple prefix code. To reduce merge index overhead, in this paper we propose a new merge mode decision algorithm which suggests using decoder derived MV candidates for merge mode decision, based upon both sum of absolute bidirectional prediction differences (SABPD) and template matching (TM) criteria. As a result, only a merge index flag is required for transmission instead of prefix coded index. The experimental results reveal that the proposed algorithm achieves up to 1.46% of bit rate reduction (with average 0.73% reduction), compared to the method used in the HEVC.**

Keywords— High efficiency video coding (HEVC); Merge mode decision

I. INTRODUCTION

The newly developed High Efficiency Video Coding (HEVC) achieves much better coding performance with 40%~60% bit rate reduction on high definition video sequences (e.g., 1080p), compared to the previous coding standard H.264/AVC. This is due to that more complicated coding techniques are employed in the HEVC. For example, the HEVC has extended inter/intra prediction with block size up to 64x64 for mode decision instead of 16x16 macroblock and used three hierarchical unit representations (including coding unit (CU), prediction unit (PU) and transform unit (TU)) to optimize the coding efficiency based on the quad-tree structure.

Like the H.264/AVC encoder that provides skip and direct mode coding in P and B frames, the HEVC also employs merge skip/2Nx2N mode in inter mode prediction in which the motion information is obtained directly from previously encoded blocks and the motion vector differences (MVD) information is not required for transmission, leading to great overhead reduction within the coded bit stream. In H.264/AVC the MV prediction (MVP) is obtained from the median of MVs of spatially adjacent blocks or co-located block from adjacent frame; while in HEVC an advanced MVP (AMVP) is proposed in which five MV candidates obtained from both spatial and temporal blocks are used for merge skip/2Nx2N mode decision [1]. As illustrated in Fig. 1, among the five MVs four candidates are selected from spatially adjacent blocks A, B, C and D (or E), and one candidate is selected from temporal blocks T_0 or T_1. The repeated MVs are eliminated and a merge candidate insertion technique is employed to find the other MVP candidates. The insertion method sequentially uses combined bi-predictive scheme, scaled bi-predictive scheme and zero MV to insert new MV candidates.

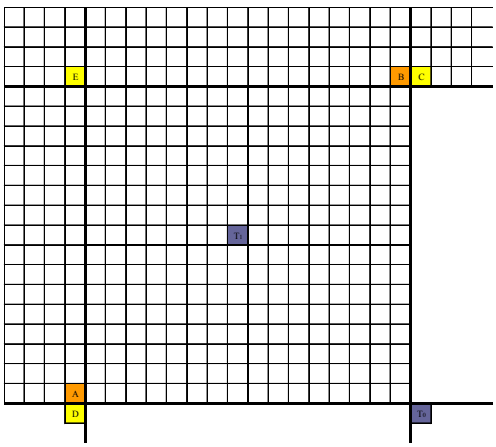

Fig. 1 MV candidates used in merge mode decision

The five MVP candidates are indicated by a merge index that is entropy-coded using a simple prefix code, with

Amazon.com, Inc., Amazon.com Services LLC - Ex. 1025, Page 000001
IPR2024-00604 (Amazon.com, Inc., Amazon.com Services LLC v. Nokia Technologies Oy)

978-1-4799-4963-2/14/$31.00 ©2014 IEEE

associated code words displayed in TABLE I. Although the spatiotemporal merge mode decision proposed in HEVC is shown to achieve better coding efficiency, compared to temporal direct mode decision or spatial direct mode decision respectively, it requires transmitting the merge index information. To reduce merge index overhead, in this paper we first investigate decoder derived MV candidates for merge mode decision and no merge index information is necessary for transmission, based upon both sum of absolute bidirectional prediction differences (SABPD) and template matching (TM) criteria. The experimental results reveal that the proposed algorithm achieves slightly worse coding performance, due to the low correct rate when using SABPD/TM criterion to choose the best MV candidate. To improve the coding efficiency, in this paper we propose a hybrid technique for merge mode decision in which we first use SABPD/TM as criterion to select the best MV candidate among merge indexes {1,2,3 and 4} and compare the best MV candidate with the candidate with merge index 0 based on the RDO technique. The one with minimum rate distortion cost is selected for merge mode decision. As a result, only a merge index flag is required for transmission, and the experimental results demonstrate that bit rate reduction can be achieved.

Merge idx	bin
0	0
1	10
2	110
3	1110
4	1111

TABLE I Associated code words for merge index

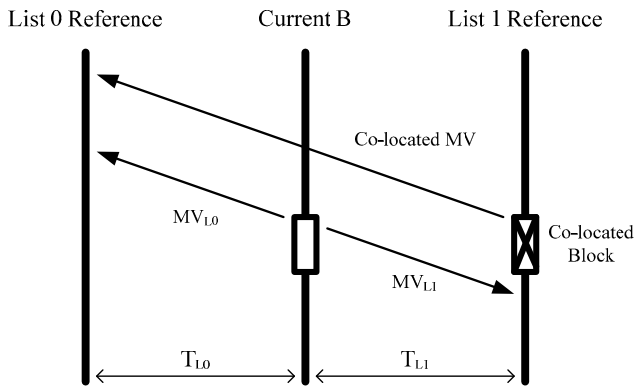

Fig. 2 Bidirectional prediction in merge mode decision

II. DECODER-DERIVED MV FOR MERGE MODE DECISION

In HEVC, the inter mode prediction for generalized B frames could use forward as well as backward reference frames (namely *List 0 reference* and *List 1 reference*) for multiple predictions. The merge mode decision uses bidirectional predictions, and the forward MV_{L0} and backward MV_{L1} are derived from the five MV candidates MV_k. As illustrated in Fig. 2, the motion vectors MV_{L0} and MV_{L1} for merge mode are calculated as

$$MV_{L0,k} = \frac{T_{L0}}{T_{L0}+T_{L1}} \cdot MV_k = (r_k, s_k) \quad (1)$$

$$MV_{L1,k} = \frac{T_{L1}}{T_{L0}+T_{L1}} \cdot MV_k = (u_k, v_k) \quad (2)$$

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 merge mode is obtained by averaging the associated blocks in these references

$$\widetilde{B}_{t,k}(i,j) = \frac{1}{2} B_{t-1,L0}(i+r_k, j+s_k) + \frac{1}{2} B_{t+1,L1}(i-u_k, j-v_k) \quad (3)$$

In the merge mode decision the HEVC encoder employs the RDO technique to choose the best MV candidate, by minimizing the rate-distortion cost function

$$J_k = SSD_k + \lambda \cdot R_k \ , \ k=0, 1, 2, 3 \text{ and } 4 \quad (4)$$

where SSD_k is the sum of the squared differences between the current block $B_t(i,j)$ and the bidirectional predicted block $\widetilde{B}_{t,k}(i,j)$ in (3) and R_k represents the truly encoded bits associated with the chosen MV candidate MV_k that includes merge index overhead, as shown in TABLE I.

Although the minimum rate distortion cost can be obtained in merge mode decision using RDO technique, the merge index overhead given in TABLE I for merge mode is required for transmission. The experimental result shows that it takes average 3% of total redundant bits for the merge index overhead and this degrades the coding efficiency.

Amazon.com, Inc., Amazon.com Services LLC - Ex. 1025, Page 000002

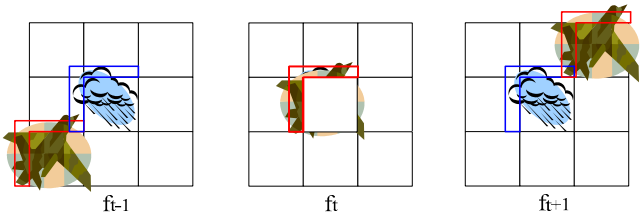

Fig. 3 Occlusion in current frame f_t

To eliminate merge index overhead, in this section we investigate a decoder MV derivation scheme to choose the best MV candidate and no overhead is required for transmission. We have observed that the difference or the prediction error between $B_t(i,j)$ and $\widetilde{B}_{t,k}(i,j)$ usually takes up much in merge mode decision when the difference between forward and backward prediction blocks $B_{t-1,L0}(i+r_k, j+s_k)$ and $B_{t+1,L1}(i-u_k, j-v_k)$ becomes large, resulting in great ghost effect. The sum of absolute bidirectional prediction differences (SABPD) between the prediction blocks can then be used as a criterion to choose the best MV candidate, and furthermore the SABPD criterion can also be derived at the decoder [2]-[3]. As a result no overhead is necessary for transmission.

This phenomenon could be not true in some cases like occlusion area etc. As illustrated in Fig. 3, the cloud is occluded by the plane in current frame f_t and it could have a smaller difference between forward and backward prediction blocks, As a result the cloud, i.e., the zero MV could be chosen as the best MV candidate for merge mode decision. To reduce this inaccuracy, we apply the decoder-derived template matching (TM), which has been widely studied for inter mode decision [4], together with SABPD to choose the best MV candidate in merge mode decision. Instead of using RDO technique given in (4) we select the best MV candidate by minimizing both SABPD/TM between prediction blocks

$$MV_{best} = \arg\min_{MV_k}[SABPD_k + TM_{b,k} + TM_{f,k}, k \in \{0,1,2,3,4\}] \quad (5)$$

with

$$SABPD_k = \sum_{i,j} \left| B_{t-1,L0}(i+r_k, j+s_k) - B_{t+1,L1}(i-u_k, j-v_k) \right|$$

$$TM_{b,k} = \sum_{x,y} \left| B_t(x,y) - B_{t-1,L0}(x+r_k, y+s_k) \right|$$

$$TM_{f,k} = \sum_{x,y} \left| B_t(x,y) - B_{t+1,L1}(x-u_k, y-v_k) \right|$$

The inverse L-shape template matching region is obtained by extending N/2 pixels from the top and left of the target region for skip/2Nx2N merge mode, as shown in Fig. 4.

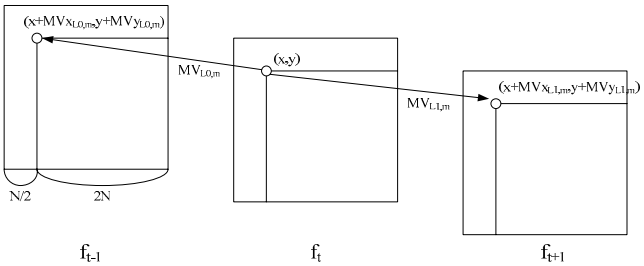

Fig. 4 Template matching

The experiment shows that the proposed decoder-derived MV method achieves slightly worse coding performance. This is due to that only 50% of best MV candidates are selected when using SABPD/TM as the criterion, leading to performance degradation although no overhead for the merge index is required for transmission.

III. HYBRID MERGE MODE DECISION FOR HEVC

The HEVC employs the RDO technique to select the best MV candidate for merge mode decision but with extra overhead required for transmission; while no overhead is necessary for transmission in the decoder-derived MV method but with inaccurate selection of best MV candidates. An intensive experiment shows that based upon RDO technique the merge index 0, with a probability as high as 0.68, is selected as the best MV for merge mode decision. To improve the coding efficiency, in this section we propose a hybrid merge mode decision for HEVC which accounts for advantages of both techniques. The proposed algorithm is depicted in Fig. 5 for illustration.

As shown, after the five MV candidates are determined we apply the SABPD/TM criterion to select the best MV candidate among merge indexes 1, 2, 3 and 4. We then use the RDO technique to choose the best MV from index 0 and the best MV candidate selected by SABPD/TM. In the proposed method, only a 1-bit index flag, indicating index 0 or the one selected from the other indexes, is necessary for transmission, as shown in TABLE II. The experiment shows that average 80% of best MV can be selected using the proposed algorithm.

Amazon.com, Inc., Amazon.com Services LLC - Ex. 1025, Page 000003

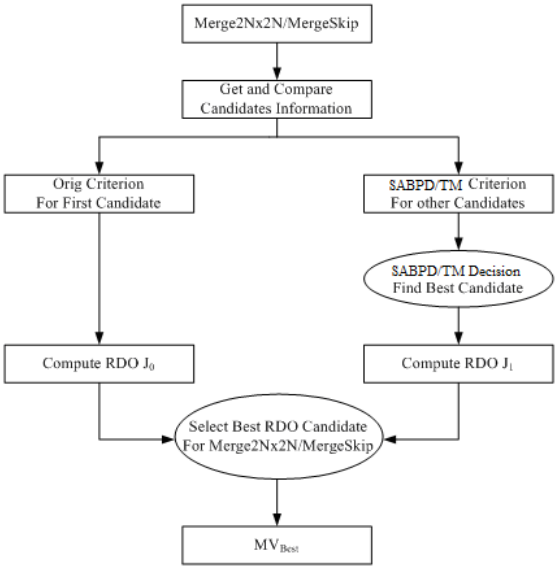

Fig. 5 Proposed merge mode decision

Merge idx	bin	
0	0	Orig Criterion
1	1	SABPD/TM Criterion
2		
3		
4		

TABLE II 1-bit index flag for merge mode decision

IV. EXPERIMENTAL RESULTS

In this section, we compare the performance of the proposed algorithm with the method proposed in HEVC. In the proposed method, based on RDO technique the best MV is selected from the index 0 and the one from the other four indexes using the SABPD/TM criterion. We implement these algorithms into the HM encoder HM4.0 to evaluate their performance. The simulation conditions are summarized in TABLE III.

Coder Version	HM4.0
Struct	Random access
GOP Struct	IBBB…
Encoding Frames	100
QP	22,27,32,37
Entropy Coding	CABAC
Test Sequence	240p,480p,720p,1080p

TABLE III Experimental settings

QP=22,27,32,37	BDPSNR(dB)	BDBR(%)

Sequence		SABPD/TM	Proposed	SABPD/TM	Proposed
1920x1080	Cactus	-0.016	0.024	0.662	-1.096
	Kimono1	-0.043	0.037	1.304	-1.150
	ParkScene	-0.040	0.020	1.222	-0.612
1280x720	vidyo1	-0.019	0.044	0.540	-1.310
	vidyo3	-0.033	0.019	0.928	-0.605
	vidyo4	-0.018	0.037	0.578	-1.280
832x480	BQMall	-0.028	0.008	0.675	-0.204
	Keiba	-0.014	-0.011	0.388	0.255
	PartyScene	-0.028	0.006	0.612	-0.120
416x240	BasketballPass	-0.015	0.033	0.300	-0.671
	Flowervase	0.019	0.087	-0.313	-1.455
	Mobisode2	-0.042	0.025	0.902	-0.533
AVG.		-0.023	0.027	0.650	-0.732

TABLE IV BDPSNR and BDBR comparisons

The performance is compared based upon Bjontegaard Delta PSNR (BDPSNR) and Delta Bit Rate (BDBR) [5] for QP=22, 27, 32 and 37. TABLE IV displays the BDPNR and BDBR results, as compared with original merge mode decision proposed in HEVC. As demonstrated, there is an average PSNR loss of 0.023 dB for SABPD/TM, as compared to the original method, or corresponding to 0.65% of bit rate increment; while the proposed algorithm can achieve average 0.03 dB PSNR gain over the original method, or corresponding to 0.7% of bit rate reduction. As shown, the proposed algorithm can save up to 1.46% of bit rate for *flowervase* sequence. The results clearly reveal the superiority of the proposed merge mode decision algorithm to the original algorithm proposed in HEVC.

V. CONCLUSION AND FUTURE WORK

In this paper we have suggested a new merge mode decision algorithm for HEVC which employs SABPD/TM criterion to select best MV candidates among indexes 1, 2, 3 and 4. The SABPD/TM criterion can be derived at the decoder. The proposed algorithm then employs the RDO technique to choose the best MV from index 0 and the best MV candidate selected by SABPD/TM, and only 1-bit index flag is required for transmission. The experimental result shows that up to 1.46% of bit rate can be saved compared to the original method proposed in HEVC. Average 0.7% of bit rate reduction can be achieved in the proposed algorithm.

This work only considers five MV candidates (as proposed in HEVC) for merge mode decision. In the future work we will consider other MV candidates from temporal domain (i.e., from the adjacent frames) and employ the SABPD/TM criterion to select the best MV candidates from these candidates to improve coding efficiency, in addition to indexes 1, 2, 3 and 4 proposed in HEVC.

Amazon.com, Inc., Amazon.com Services LLC - Ex. 1025, Page 000004

REFERENCES

[1] CT-VC, "Unification of derivation process for merge mode and MVP," JCTVC-F419, JCT-VC Meeting, Torino, Jul. 2011.

[2] A. M. Huang and T. Nguyen, "Correlation-based motion vector processing with adaptive interpolation scheme for motion-compensated frame interpolation," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 740-752, April 2009.

[3] Y. Ling, J Wang, Y. Liu and W. Zhang, "Spatial and temporal correlation based frame rate up-conversion," in *Proc. IEEE ICIP*, pp. 909-913, Sept. 2008.

[4] S. Kamp, M. Evertz and M. Wien, "Decoder side motion vector derivation for inter frame video coding," in *Proc. IEEE ICIP*, pp. 1120-1123, 2008.

[5] G. Bjontegaard, "Calculation of average PSNR difference between RD curves", ITU-T Q.6/16, Doc. VCEG-M33, Apr. 2001.

Amazon.com, Inc., Amazon.com Services LLC - Ex. 1025, Page 000005