A MOTION AIDED MERGE MODE FOR HEVC

*Hao Li, Kui Fan, Ronggang Wang**, *Ge Li, Wenmin Wang*

School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University
lihao@sz.pku.edu.cn, kuifan@pku.edu.cn, rgwang@pkusz.edu.cn, gli@pkusz.edu.cn, wangwm@ece.pku.edu.cn

ABSTRACT

Merge prediction is a practical inter-technique in HEVC, which can significantly improve the coding efficiency, especially for homogeneous regions in video sequences. In this paper, a motion aided merge mode (MAMM) is proposed to achieve a better trade-off between the prediction accuracy and bit rate. Different from the traditional merge mode in HEVC, MAMM is accomplished by a small motion obtained by searching in a specific search region. The search range is comprised of a number of points with high occurrence possibilities. The motion vector difference (MVD) is coded by Huffman coding in MAMM and the Huffman coding table is generated according to the statistical frequency of each possible MVD value. The proposed method is implemented on top of the HEVC reference software (HM-16.15), and experimental results show that 0.6% BD-rate reduction is achieved under Random Access (RA) configuration.

Index Terms— HEVC/H.265, Inter-Prediction, Merge Mode, Huffman Coding, Video Coding

1. INTRODUCTION

The current high efficiency video coding (HEVC) [1] standard with its first version published as ITU-T H.265 and ISO/IEC 23008-2 in 2013 is a state-of-the-art video coding standard, which achieves up to 50% bit rate reduction at the same subjective quality of its predecessor, the H.264/AVC standard [2]. The block-based hybrid coding scheme of HEVC consists of intra and inter prediction, transform, quantization, CABAC, and in-loop filter. Larger coding blocks (up to 64x64) and more block partitions are supported for coding, prediction, and transform processes in HEVC, and bring considerably higher compression efficiency. In inter prediction, several efficient coding tools, such as merge mode [3] and advanced motion vector prediction (AMVP), are included in HEVC to further reduce temporal redundancy.

In HEVC, merge mode mainly consists of two steps. The first step is to construct the merge candidate list. Each merge candidate includes the prediction direction, the motion vector and the corresponding reference index. The second step is to select the best merge candidate by minimizing the error between the merge prediction block and the original block. In contrast to other inter prediction modes with motion estimation (ME) process, motion information in merge mode are derived by copying from spatially neighboring blocks or temporally collocated blocks. Only a merge flag and a merge index need to be transmitted to decoder. Consequently, the bit consumption for motion parameters in merge mode is significantly decreased.

Some methods are proposed to improve merge mode. In [4], Chien et al. use picture order count (POC) instead of reference index when removing redundant merge candidates, and the performance of merge mode is improved. In [5], the partition-adaptive merge candidate derivation is proposed to refine the merge candidate list. Specifically, the merge candidate derived from the same hierarchical coding unit (CU) is excluded from the merge candidate list of the current CU. In order to represent the complex motion more accurately, Zhang et al. [6] propose a new merge mode by introducing bilinear interpolation model, six-parameter affine model, and four-parameter affine model. In the scalability extension of HEVC (SHVC), HoangVan et al. [7] propose an improved enhancement layer merge mode prediction solution that adaptively refines the motion vector candidates.

In this paper, we present another improvement of merge mode, i.e. motion aided merge mode (MAMM). In MAMM, the merge candidate is enhanced by a simple ME in encoder. During the process of ME, we studied the occurrence possibility of each search point within a defined region, and a specific search process is designed. In order to reduce the corresponding bit rate, the motion vector difference (MVD) is coded by means of Huffman coding [8]. MAMM is integrated into the HEVC reference software (HM-16.15), and experimental results show that on average 0.6% BD-rate reduction can be achieved.

This project was partly supported by National Natural Science Foundation of China 61672063, Shenzhen Peacock Plan, Shenzhen Research Projects of JCYJ20160506172227337 and GGFW2017041215130858, and Guangdong Province Projects of 2014B090910001.

The remainder of this paper is organized as follows. Section 2 provides the details of the proposed algorithm. Section 3 gives the experimental results and Section 4 concludes the paper.

2. MOTION AIDED MERGE MODE

In the video coding, the accuracy of motion information is very important for effectively reducing the temporal redundancy. In general, the temporal motion information is obtained through the ME process. Although a high prediction accuracy can be achieved, the computational complexity of ME is very high. In addition, the bit overhead is large. Since many syntax elements need to be signaled, such as motion vector difference, reference index and so on. For saving both the computation and the bit consumption of motion information, merge mode is applied in HEVC. The motion information in merge mode is just derived from the previously coded blocks. As a result, merge mode cannot provide precise prediction results, especially in the areas with complicated movements. However, it is very efficient for homogeneous regions and still areas. In order to better balance the coding bit rate and the accuracy of the prediction, MAMM is presented as an enhanced merge mode.

2.1. Overview of MAMM

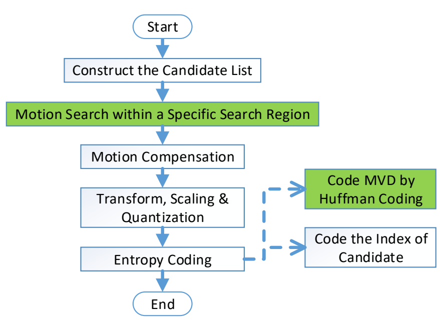


Fig. 1. Flow chart of the proposed MAMM.

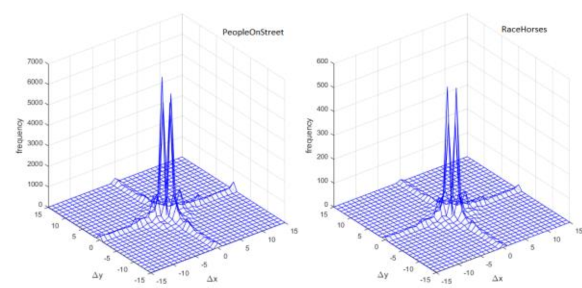
Fig. 1 shows the flow chart of MAMM. In particular, the differences between MAMM and merge mode are marked with the green blocks in Fig. 1. At first, the candidate list of the traditional merge mode in HEVC is constructed, and it includes five candidates at most. Each candidate consists of motion vector, reference index, and prediction direction. Secondly, in order to provide precise prediction block, a simple motion search is applied based on merge candidates. After that, motion compensation, transform, quantization, and entropy coding are performed. The proposed MAMM method is assessed by rate distortion optimization (RDO). A CU-level flag is used to indicate whether MAMM is applied in the current CU. If MAMM is selected, some additional information, including merge index and MVD, is transmitted to decoder. Besides, because the traditional merge mode is very effective, the proposed MAMM is applied only if the traditional merge mode is not selected.

The inter-coded CUs can be partitioned into prediction units (PUs) in two ways, that is, symmetric motion partitioning and asymmetric motion partitioning (AMP) [9]. The proposed MAMM is applied to all inter partitions except for NxN. Considering the high computational complexity of the ME process, a small search region is specified, and the search method is simplified for CUs with multiple PUs. The details are described specifically in Section 2.2.

In entropy coding of MAMM, in addition to the index of the candidate, MVD is also signaled in bit streams. In order to reduce the bit rate, we use the Huffman coding method, in which two components of MVD are combined and signaled in bit streams. In our experiments, this method is proved to cost fewer bits than coding two components separately. The bit signaling method of MVD for MAMM is depicted in Section 2.3.

2.2. Specific Search Region

Taking both the cases of the slow movement and the complexity of motion search into consideration, it is necessary to refine the motion vectors in a small search region. In the proposed method, we constrain the search region in a small specified area, and refine it according to the statistical frequency of each search point.



(a) (b)

Fig. 2. Statistical MVD frequency of two sequences (a) PeopleOnStreet (b) RaceHorses.

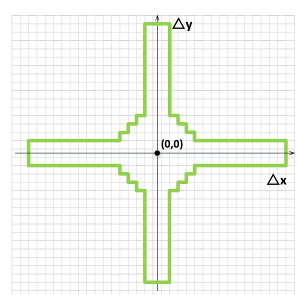


Fig. 3. The specific search region.

Firstly, we exploit the possibility of each search point to be the best motion vector by searching a 30x30 square region surrounding the initial searching point. In the proposed method, the MV of merge candidate is set as the initial searching point. Fig. 2 demonstrates the possibilities of MVD values to be selected on two sequences ("PeopleOnStreet", "RaceHorses"). It can be observed that the search points with high frequencies are located in the central area of the square region, and the search points along two axes also have a high possibility to be selected. According to the statistical results, it can be found that a small amount of search points play dominant roles, and other points are scarcely selected.

Based on the above analysis, we design a small and specific region only including the high-frequency search points around the initial search point, as shown in Fig. 3 (green line). The specific region can be expressed as:

$$\Omega_{Region} = \begin{cases} |\Delta y| \le 15 & , \text{ if } |\Delta x| \le 1 \\ |\Delta y| \le 6 - |\Delta x| & , \text{ if } 1 < |\Delta x| \le 4 \\ |\Delta y| \le 1 & , \text{ if } 4 < |\Delta x| \le 15. \end{cases} \tag{1}$$

where Ω_{Region} represents the specific search region and each MVD is represented by a pair of integer variables (Δx, Δy).

Considering the high correlation between the MVs of two PUs in a CU, after the MV of the first PU is obtained in MAMM, the motion search of another PU is simplified. To clarify, for the first PU, the integer MV is obtained by the traditional fast search algorithm. In the sub-pixel motion search, eight 1/2-pel positions surrounding the integer MV and eight 1/4-pel positions surrounding the best 1/2-pel positions are checked to get the best fractional MV. After that, for the second PU, only four adjacent positions around the best MV of the first PU are checked to obtain the best MV.

Compared with the process of ME in HEVC, MAMM has advantages of lower complexity. Furthermore, MAMM can provide more precise prediction results than the traditional merge mode.

2.3. Huffman Coding

The Exp-Golomb coding is a kind of variable-length code that can code an arbitrary non-negative integer. In HEVC, both the horizontal and the vertical components of MVD are coded by Exp-Golomb coding. Nevertheless, the components of MVD in MAMM are relatively small. Due to the fact that Exp-Golomb code cannot deal with small values efficiently and may lead to additional bit overhead, we exploit a Huffman based coding method to code MVD in MAMM.

Table 1. A part of Huffman Code Table for MAMM

MVD	Code word	MVD	Code word	MVD	Code word
(0,-1)	010	(0, 3)	011001	(-6, 0)	1110011
(0, 1)	001	(0,-2)	011000	(6, 0)	1000110
(-1, 0)	110	(-1, 1)	000111	(0,-7)	1000100
(1, 0)	101	(-1,-1)	000110	(0, 7)	1000010
(-2, 0)	11101	(1, 1)	000001	(-2, 1)	01111110
(2, 0)	10010	(-3, 0)	111110	……	
(0,-3)	011110	(0, 2)	111100	……	
(1,-1)	011100	(3, 0)	100000	……	

In order to minimize the bits of coding MVD, the two components of MVD are combined to be coded. In other words, each pair of MVD corresponds to one codeword. A part of Huffman code table for MVD of MAMM is shown in Table I, and the whole code table has stored in both encoder and decoder. As a result, the encoder can obtain the codeword of the corresponding MVD and transmit it to the decoder. While, the MVD can be decoded by querying this table after codeword is obtained at the decoder.

The Huffman code table is generated by calculating the probability of each MVD. In addition, the occurrence times of each MVD is collected by encoding video sequences using HEVC encoder.

In our experiments, the proposed Huffman based coding method is obviously superior to the Exp-Golomb coding method.

3. EXPERIMENTAL RESULTS

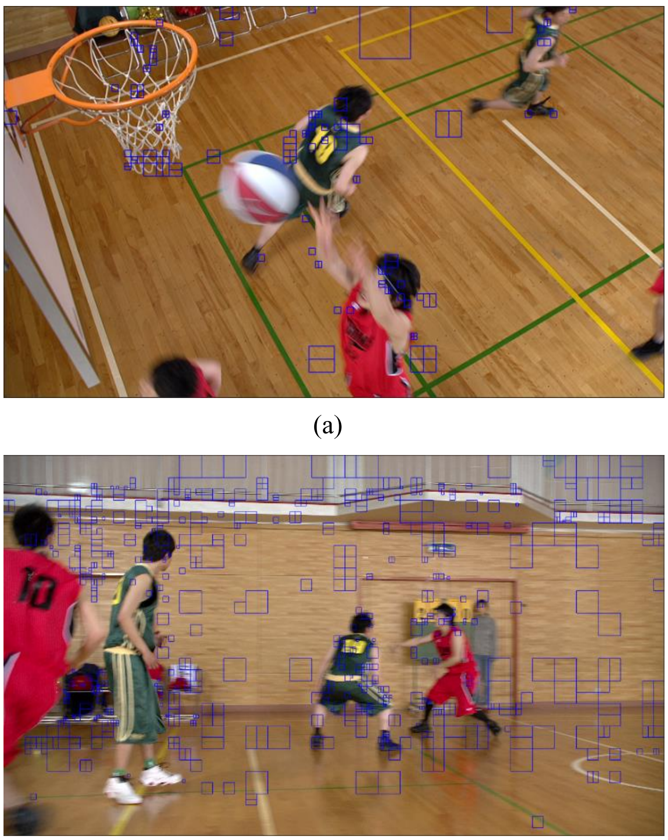
To evaluate the coding efficiency, the proposed method has been integrated into the HEVC reference software (HM-16.15). Experiments are carried out under Random Access (RA) coding configuration, and the common test condition of HEVC [10] is followed. Five natural video classes (A to E) are tested. Class A to E represent 4K, 1080p, WVGA, QWVGA, and 720p sequences, respectively. Bjøntegaard delta bitrate (BD-rate) with piece-wise cubic interpolation method is used to assess the objective coding performance of the proposed motion aided merge mode.

Table 2. BD-rate of the proposed MAMM

Class	Sequence	Y	Cb	Cr
Class A	Traffic	-0.2%	-0.1%	0.0%
	PeopleOnStreet	-0.3%	-0.3%	-0.1%
	Nebuta	-0.7%	0.3%	-0.1%
	SteamLocomotive	-0.9%	-0.2%	-0.4%
	Average	**-0.6%**	**0.0%**	**-0.2%**
Class B	Kimono	-0.4%	-0.3%	-0.2%
	ParkScene	-0.3%	-0.1%	-0.1%
	Cactus	-1.0%	-0.8%	-0.8%
	BasketballDrive	-1.5%	-0.7%	-1.1%
	BQTerrace	-0.8%	-0.3%	-0.1%
	Average	**-0.8%**	**-0.5%**	**-0.5%**
Class C	BasketballDrill	-0.3%	-0.2%	-0.4%
	BQMall	-0.5%	0.0%	-0.4%
	PartyScene	-0.5%	-0.3%	-0.4%
	RaceHorsesC	-0.6%	-0.6%	-0.6%
	Average	**-0.5%**	**-0.3%**	**-0.5%**
Class D	BasketballPass	-0.6%	-0.4%	-0.8%
	BQSquare	-0.5%	-0.2%	-0.1%
	BlowingBubbles	-0.3%	-0.3%	-0.3%
	RaceHorses	-0.5%	-0.6%	-0.9%
	Average	**-0.5%**	**-0.4%**	**-0.5%**
Class E	FourPeople	-0.4%	-0.2%	-0.1%
	Johnny	-0.7%	-0.4%	-0.2%
	KristenAndSara	-0.5%	-0.2%	-0.2%
	Average	**-0.5%**	**-0.3%**	**-0.2%**
All		**-0.6%**	**-0.3%**	**-0.4%**

The overall experimental results are summarized in Table II. As showed, the BD-rate gain is 0.6% for luma component on average, and 0.3% and 0.4% for two chroma components in RA case. The decoding time of proposed MAMM keeps almost unchanged and the encoding time is increased by 40%~50%.

Fig. 4 visualizes the blocks applied MAMM in the 13th frames of BasketballDrill (WVGA) and BasketballDrive (1080p) at QP 22. For BasketballDrill sequence, the background is stationary while the players and the basketball are moving. In contrast, the background of BasketballDrive sequence is smoothly moving all the time. In the second sequence, the number of blocks applied MAMM is far larger than that of the first sequence. It is implied that the blocks applied MAMM are mostly located at the areas with small-range movements, such as the slowly moving background and objects.



(a)

(b)

Fig. 4. Blocks applied MAMM marked with blue boxes in the 13th frames of (a) BasketballDrill (WVGA) (b) BasketballDrive (1080p) sequence.

The compression performance of the proposed algorithm depends on the characteristics of sequences. Empirically, the greater area of the region with slow motion, the better the coding performance.

4. CONCLUSION

In this paper, a motion aided merge mode for inter pictures is presented. To reduce complexity, a specific search region is subtly designed according to the statistical frequency. The MV in merge candidate is searched within that region. MVD is efficiently coded by a Huffman based coding method. Experimental results show the proposed method is able to improve the coding performance in RA configuration, especially for the sequences with the background motion derived by the camera movement.

5. REFERENCES

[1] G. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[2] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[3] P. Helle, S. Oudin, B. Bross, et al, "Block Merging for Quadtree-Based Partitioning in HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1720–1731, Dec. 2012.

[4] W.-J. Chien, J. Chen, et al, "Modification of merge candidate derivation," *ITU-T/ISO/IEC Joint Video Exploration Team (JVET)* document JVET-B0058, Feb. 2016.

[5] S. Iwamura, K. Iguchi, et al, "Partition-adaptive merge candidate derivation," *ITU-T/ISO/IEC Joint Video Exploration Team (JVET)* document JVET-D0107, Oct. 2016.

[6] N. Zhang, X. Fan, D. Zhao, and W. Gao, "Motion Vector Derivation of Deformable Block," *IEEE International Conference on Image Processing*, pp. 1549-1552, Sept. 2012.

[7] X. HoangVan, J. Ascenso, and F. Pereira, "Improving enhancement layer merge mode for HEVC scalable extension," *Picture Coding Symposium (PCS)*, pp. 15–19, Jun. 2015.

[8] D. Huffman, "A Method for the Construction of Minimum Redundancy Codes", *Proceedings of the I.R.E.*, pp. 1098-1101, 1952.

[9] I.-K. Kim, S. Lee, M.-S. Cheon, T. Lee, and J. Park, "Coding efficiency improvement of HEVC using asymmetric motion partitioning," *Broadband Multimedia Systems and Broadcasting (BMSB), 2012 IEEE International Symposium on*, Seoul, pp. 1–4, Jun 2012.

[10] F. Bossen, "Common test conditions and software reference configurations," *ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC)* document JCTVC-L1100, Jan. 2013.