Adaptive Motion Vector Resolution Scheme for Enhanced Video Coding

Zhao Wang*, Jian Zhang*, Nan Zhang+, Siwei Ma*

* *Institute of Digital Media & Cooperative Medianet Innovation Center, Peking University Beijing, 100871, China*
{zhaowang, jian.zhang, swma}@pku.edu.cn

+ *School of Biomedical Engineering, Capital Medical University Beijing, 100069, China*
zhangnan@ccum.edu.cn

***Abstract**:* In the state-of-the-art H.265/HEVC video coding standard, the motion vector is always fixed to be 1/4-pixel resolution for the entire video sequence regardless of the different video contents, which is not efficient for prediction coding. In this paper, we propose a frame level adaptive motion vector resolution selection scheme based on a rate-distortion model in terms of motion vector resolution. In the proposed rate-distortion model, the relationship between the distortion and the motion vector resolution is approximated with a linear model. And a rate model of motion vector is built, which reflects the relationship between the coding bits of motion vector and its value. With the proposed rate-distortion model, an optimal motion vector resolution minimizing the total rate-distortion cost will be selected for each frame. Experimental results show that the proposed scheme can achieve 1.5%, 1.3% and 2.5% BD-rate gain on average for Random Access, Lowdelay-B and Lowdelay-P configurations without complexity increment.

1. Introduction

Block-based motion compensation is widely used in video coding because of its capability to reduce temporal redundancy between consecutive frames. In the earliest video coding standard H.261, integer-pel motion vector (MV) resolution is adopted, where only prediction at full pixel locations can be obtained. However, the motion between consecutive frames is not necessarily integer pixel. Therefore, subpel MV with half-pel resolution has been introduced into MPEG-2 and H.263, which significantly improves the coding efficiency. Later, quarter-pel MV resolution is adopted in H.264/AVC [1], and is also adopted in the latest H.265/HEVC [2]. Usually the motion vectors with higher resolution can provide better prediction accuracy, however the higher resolution motion vectors also require more bits to be coded. During the development of HEVC, eighth-pel MV resolution has been proposed but it is not adopted because no more coding gain can be achieved compared with quarter-pel resolution [3]. Hence, the trade-off between the coding efficiency and the MV resolution has been an important research issue in block-based motion compensation video coding.

In [4], Girod first gives a theoretical analysis for the coding gain of increasing the MV resolution. A practical adaptive MV resolution scheme is proposed by J. Ribas-Corber in [5-6]. In [5], the authors analytically obtain the rate model of residual and the motion vectors in terms of the MV resolution and several other parameters related to the texture complexity and motion activity. These models show that higher MV resolution should be used for the area with more texture complex and vice versa. In [7], the texture feature is also used to predict the potential coding gain while increasing the MV resolution, and

1068-0314/16 $31.00 © 2016 IEEE
DOI 10.1109/DCC.2016.90

used to decide whether to increase the MV resolution or not. Furthermore, a surface model of the local matching error is later proposed to predict the optimal MV resolution in [8]. Moreover, it is found that the optimal MV resolution in different inter prediction pictures, e.g. B and P pictures, may be different, where lower resolution is preferable in B pictures [9].

In the recent literature [10], a progressive MV resolution scheme is proposed, where higher MV resolution is employed for the searched blocks near to the motion vector predictor (MVP) and lower resolution is employed for blocks far from the MVP. In [11], the authors analyze the potential influencing factors when select the optimal MV resolution, including the texture complexity, motion scale, inter-frame noise and quantization parameter. In the latest HM-KTA2.0[12] which is the reference software during the meeting of VCEG, an adaptive coding unit (CU) level MV resolution scheme is integrated, where the MV resolution can either quarter-pel or integer-pel, and a flag of selected MV resolution is signaled for each CU. However, the overhead at CU level limits the potential coding gain and increases the complexity in hardware design.

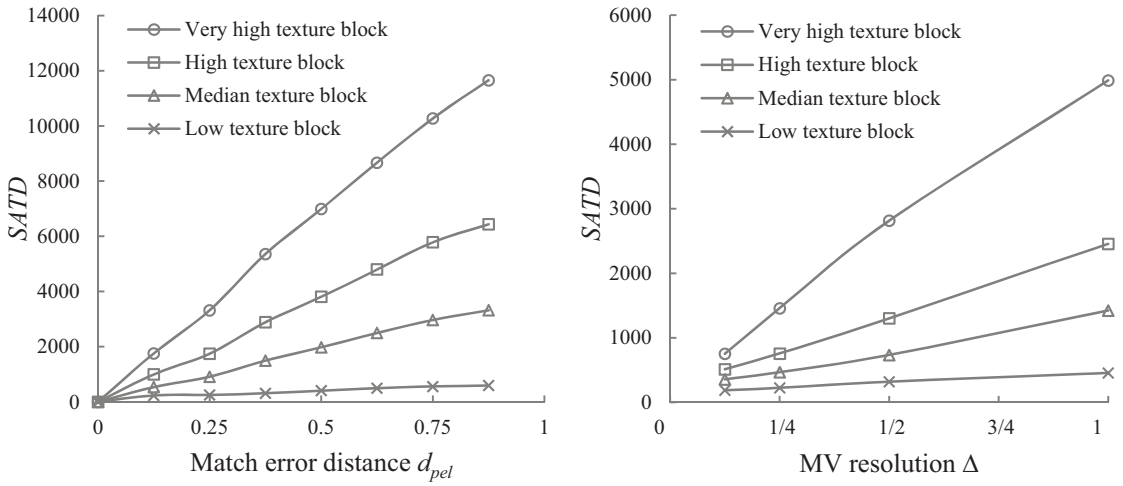
In this paper, we propose a frame level adaptive MV resolution selection scheme based on a rate-distortion model in terms of the MV resolution. For the distortion model, it is observed that there is approximately a linear relationship between the prediction distortion and the MV resolution used. Then, we analyze the distribution of MVD and build the corresponding rate model by distinguishing MVDs into three types according to their numerical values. Based on the proposed rate-distortion model, the optimal MV resolution in terms of minimizing the rate-distortion cost is selected for each frame.

The remaining of this paper is organized as follows: Section 2 and Section 3 present the derived distortion and rate model respectively. Section 4 provides the adaptive MV resolution selection scheme and some implementation issues. Section 5 shows the experimental results and Section 6 concludes this paper.

2. The Prediction Distortion Model

In this section, we derive a distortion model between the prediction residual and the MV resolution, to estimate the decrement of prediction distortion when increases the MV resolution. If increasing the MV resolution provides rich improvement of the prediction accuracy, higher MV resolution should be used to reduce the prediction distortion; otherwise, if only slight improvement of prediction accuracy is obtained, lower MV resolution is preferable to reduce the number of required coding bits.

As we know, for an ideal perfect motion compensation, the prediction distortion is reduced to zero, while the prediction block and the current block match exactly. And if the search block deviates from the perfect match point, the prediction block between the search block and the current block will increase progressively as the result of mismatch. Actually, the match error is usually not zero even at the optimal match point because of the change of pixels in the consecutive frames. This change mainly comes from quantization noise, lighting change, or from the change of object itself. Here we would first ignore these inter-frame noises, but focus on the perfect match case. For one block, we interpolate the subpel position pixels by the DCT-IF method [3], and take the subpel blocks as the prediction blocks. The distortions between the current block and each prediction block located at subpel positions are plotted in the following Figure 1-(a).



(a) No noise condition (b) Real noisy coding environment

Figure 1: Subpel MV resolution impact on block residual.

Figure 1-(a) shows how the prediction distortion increases as the searched block deviates from the current block for blocks with different texture complexity. The y-axis is the prediction distortion, which is measured by the sum of absolute transformed difference (SATD). The *x*-axis is the matching error distance d_{pel} between the searched block and the ideal matching point (namely the current block itself). For example, $d_{pel} = 0.5$ means that the searched block locates at half pel position. According to Figure 1-(a), it can be found that the prediction distortion is proportional to the distance d_{pel} between the searched block and the current block. When d_{pel} is zero, it achieves perfect match because the searched block is just the current block itself. As the matching error distance d_{pel} increases, the distortion between the searched block and the current block also increases progressively.

In the above experiment, the effect of noise is not considered. However, for the real video coding implementation, noise can't be neglected. Noise may be introduced by two primary sources. The first one is the quantization noise and the second one is some external factors such as the lighting change, camera movement and object change, etc. All of these noises can be generally considered as the match error leading to non-perfect motion compensation (MC) even for the optimal prediction block searched from the reference frames. To explore the distortion model in the noisy coding environment, we plot the prediction distortions when these four blocks are encoded with different MV resolutions in a practical codec, as shown in Figure 1-(b). The results show that the prediction distortion is proportional to the MV resolution used, which is similar to the noise-free condition.

From above experiments, it is observed that there is a linear relationship between the prediction distortion and the MV resolution both in the noise-free condition and the real coding environment. Though the distortions of all blocks decrease as the MV resolution improves, the slopes are quite different for the blocks with various texture. Usually, the blocks with rich texture benefit much from the subpel MC process, and the blocks with low texture benefit less, because only slight improvement of prediction can be obtained when increases the MV resolution. Hence, it is inferred that the slope of distortion decrease is determined by the block's texture, and the noise mainly serves as a constant which leads to non-perfect match. Thus, we model the prediction distortion *D* as:

$$D = a \cdot \Delta + N, \tag{1}$$

where Δ is the MV resolution. a is the slope determined by the block's texture and N represents the inter-frame noise.

In order to quantify the impact of block's texture on slope a, we conduct simulations for 200 blocks in each sequences. The results of sequences *BasketballPass*, *BQSquare* and *RaceHorses* are illustrated in Figure 2, where the *y*-axis is the slope a of each $H \times W$ block and the *x*-axis is the block's texture T which only considers the horizontal direction for simplification.

$$T = \sum_{i=1}^{H} \sum_{j=1}^{W} |\boldsymbol{s}(i,j) - \boldsymbol{s}(i+1,j)|, \tag{2}$$

where $\boldsymbol{s}(i, j)$ is the luma value of pixel located coordinates (i, j) and $\boldsymbol{s}(i+1, j)$ represents its right neighboring pixel. According to Figure 2, the slope a can be modeled as

$$a = \alpha \cdot T, \tag{3}$$

where α is approximately equal to 2.4 in our experiments.

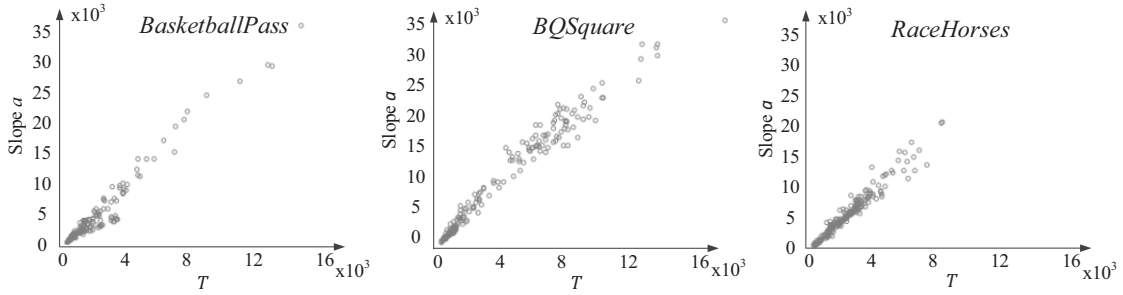


Figure 2: The impact of block's texture on the slope a.

3. The Rate Model of Motion Vector

In this section, we develop a rate model of motion vector which reflects the relationship between the number of coding bits and the magnitude of the MVs. Usually, the MV of the current block is correlated with the MVs of the neighboring blocks in the current frame or in the earlier encoded frames. Therefore, the motion vector prediction method is proposed where the MVs in neighboring blocks serve as predictors to reduce the size of the current MV. By this method, the signaling of MV is converted into transmitting the index of MVP and the MVD between the motion vector of the current block and the MVP. Considering the MVP is less related to the MV resolution, here we mainly focus on the entropy coding of MVD.

As we know, the motion parameters are coded by context-based adaptive binary arithmetic coding (CABAC) in HEVC/H.265 [13], where it uses the statistical properties to compress data such that the number of bits used to represent the data is related to the probability of the data [14]. In general, small MVDs need less bits to be coded and large MVDs need more bits. And the relationship between the coding bits of MVDs and their values are illustrated in the following Figure 3.

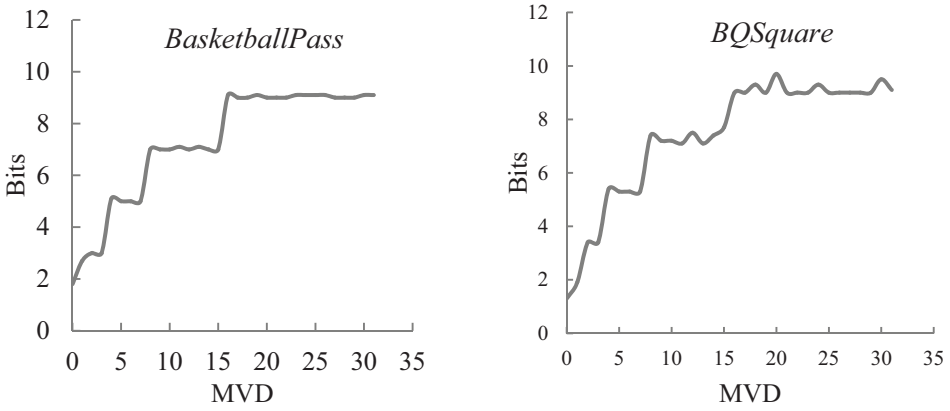
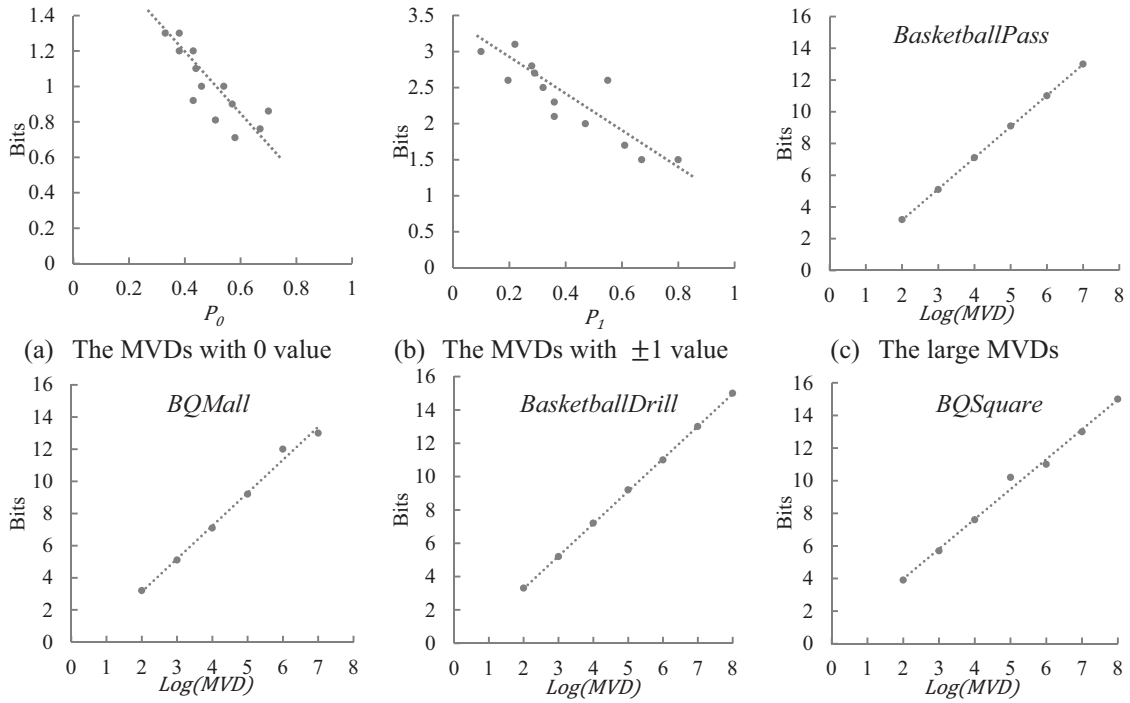


Figure 3: The relationship between MVDs' coding bits and their absolute values.

From Figure 3, it can be seen that the coding bits of MVDs increase as their values increase for the small MVDs. However, for the large MVDs, the coding bits is a staircase function, where an increase in MVD's value by a factor of 2 leads that it requires 2 more bits to be coded. Hence, it is inferred that the coding bits of small MVDs are related to their probability, while the coding bits of large MVDs are mainly determined by their values. Actually, the large MVDs take minority and the probabilities are so little that it performs no impact on the coding bits. Based on this characteristic, we classify MVDs into three types according to their absolute values: 0, ± 1 and others. For the small MVDs, the relationships between the coding bits and the probability are illustrated in Figure 4-(a) and Figure 4-(b), respectively. For the large MVDs, the rate models are plotted from Figure 4-(c) to Figure 4-(f).



(a) The MVDs with 0 value (b) The MVDs with ± 1 value (c) The large MVDs

(d) The large MVDs (e) The large MVDs (f) The large MVDs

Figure 4: The illustration of MVD's rate model

From the Figure 4-(a) and Figure 4-(b), it is observed that the coding bits of small MVDs are inverse proportional to the probabilities, which are computed by

$$P_0 = \frac{num(MVD = 0)}{num(All\_MVD)}\ ; \quad P_1 = \frac{num(abs(MVD) = 1)}{num(abs(MVD) > 0)}, \tag{4}$$

where $num()$ and $abs()$ are the count and absolute value function respectively. For the large MVDs, it can be found that there is a linear relationship between the coding bits and the logarithm of MVDs' values. Hence, we model the rate of MVD as

$$R_{MVD} = \begin{cases} a_0 \cdot P_0 + b_0 & when\ abs(MVD) = 0; \\ a_1 \cdot P_1 + b_1 & when\ abs(MVD) = 1; \\ a_2 \cdot \log_2 abs(MVD) + b_2 & when\ abs(MVD) > 1. \end{cases} \tag{5}$$

In our experiments, the parameters (a_0, b_0), (a_1, b_1) and (a_2, b_2) in above rate model (5) are equal to (-1.78, 1.82), (-2.5, -3.46) and (2, 1.2), respectively.

4. The Adaptive Motion Vector Resolution Scheme

In this section, we propose an adaptive motion vector resolution (AMVR) scheme that selects the MV resolution minimizing the rate-distortion cost for each frame, based on the above rate-distortion model in Section 2 and Section 3. As shown in Figure 5, the motion vector resolution Δ which minimizes the cost J_{RD} will be selected as the best motion resolution, as follows [15-16],

$$J_{RD}(\Delta) = D(\Delta) + \lambda R(\Delta), \tag{6}$$

where λ is the Lagrangian multiplier related to the quantization parameter (QP), D is the distortion and R is the number of bits needed for coding motion information.

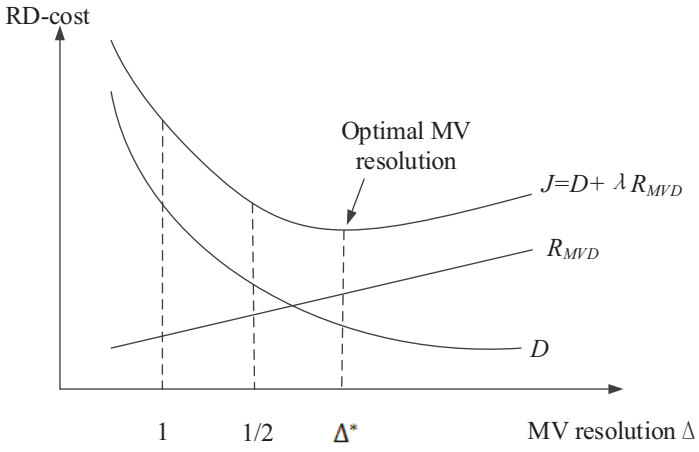


Figure 5: The illustration of the proposed adaptive MV resolution scheme.

To estimate the rate-distortion cost in different MV resolution for the current frame, only the inter mode blocks should be considered. As the prediction modes for current frame are not available before the whole frame is fully encoded, we utilize the prediction modes of the previous frame in the same temporal layer to estimate the rate-distortion cost. Let Δ and Δ' denote the MV resolution in the previous frame and the MV resolution to be used

in the current frame respectively, and let B denote all the inter blocks in the previous frame. Then the rate-distortion cost will be

$$J(\Delta') = \sum_B D(\Delta') + \sum_B R_{MVD}. \tag{7}$$

Based on the proposed distortion model (1) and the rate model (5), the $J(\Delta')$ can be expanded as

$$J(\Delta') = a \cdot (\sum_B T) \cdot \Delta' + \sum_B N + \sum_B R_{MVD}\left(\frac{\Delta}{\Delta'} \cdot MVD\right). \tag{8}$$

Because the noise term N serves as a constant which is not related to the MV resolution, there is no need to compute the value of noise. Hence, the optimal MV resolution Δ^* will be selected as follows

$$\Delta^* = Min||a \cdot (\sum_B T) \cdot \Delta' + \sum_B R_{MVD}\left(\frac{\Delta}{\Delta'} \cdot MVD\right)||$$

In the real implementation, the index of the selected resolution is coded using 2-bits fixed- length method, and the overheads on frame level can be ignored. For the start frames in a sequences, it will use the default quarter-pixel resolution.

Because the MV resolution may be different between consecutive frames, the MVPs need to be rounded to the current resolution. For example, if the MVP is at eighth subpel position and the selected MV resolution of the current frame is quarter-pixel, we will round it to the nearest quarter subpel position. For the merge mode, it still uses the original resolution of each merge candidates.

5. Experimental Results

The proposed adaptive motion vector resolution scheme is integrated into HEVC reference software HM16.2. Firstly, we verify the accuracy of the estimation model of residual and the rate of MVD. Figure 6-(a) shows the real prediction distortion and the estimated distortion of the inter blocks in each frame, and Figure 6-(b) shows the coded bits and the estimated bits of MVDs in each frame. It proves that our models perform well.

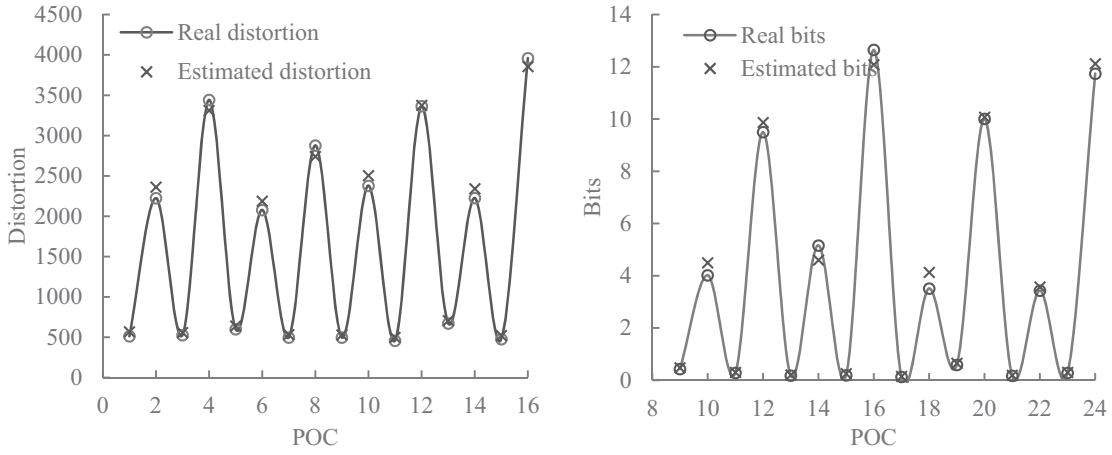


Figure 6: The real and estimated distortion/rate respectively

Table 1: The performance of proposed AMVR scheme compared with HM16.2

Class	Sequence	Random Access(RA)			Lowdelay-B (LDB)			Lowdelay-P (LDP)		
		Y	U	V	Y	U	V	Y	U	V
A	Traffic	-1.3	-1.2	-1.7	-1.2	-0.8	-1.1	-0.7	-0.5	-0.9
	PeopleOnStreet	-0.7	-1.2	-1.3	-0.9	-1.1	-0.9	-1.1	-1.4	-1.9
B	Kimono	-1.8	-2.0	-2.5	-1.4	-1.6	-1.7	-0.8	-1.1	-1.2
	ParkScene	-0.3	-0.3	-0.4	-0.4	-0.5	-0.6	-0.5	-0.6	-0.9
	Cactus	-0.7	-0.4	-0.5	-0.7	-0.7	-0.9	-0.6	-0.5	-0.7
	BasketballDrive	-1.7	-1.9	-2.0	-1.3	-1.2	-1.4	-0.9	-1.2	-1.7
	BQTerrace	-1.3	-1.5	-1.5	-1.1	-1.4	-1.2	-3.8	-3.8	-4.3
C	BasketballDrill	-1.9	-2.7	-2.6	-1.7	-2.3	-2.1	-2.3	-4.4	-3.6
	BQMall	-1.4	-0.5	-1.2	-1.1	-0.7	-0.9	-2.3	-2.5	-2.0
	PartyScene	-2.8	-3.4	-2.9	-2.1	-2.7	-3.2	-7.2	-6.3	-6.5
	RaceHorsesC	-0.7	-0.8	-0.9	-0.5	-0.6	-0.7	-0.5	-0.1	-0.3
D	BasketballPass	-0.8	-1.6	-1.1	-0.7	-1.1	-1.3	-0.8	-0.4	-0.4
	BQSquare	-6.7	-4.7	-3.2	-5.4	-4.8	-3.8	-17.3	-12.9	-14.8
	BlowingBubbles	-1.3	-1.1	-1.6	-1.6	-1.7	-2.3	-2.6	-2.9	-1.6
	RaceHorses	-1.2	-1.5	-2.3	-1.0	-1.3	-1.4	-0.6	-0.5	-0.7
E	FourPeople	-1.0	-0.4	-0.5	-0.7	-0.6	-1.1	-0.7	-0.8	-0.6
	Johnny	-0.8	-1.1	-1.4	-1.1	-1.2	-0.9	-1.4	-1.7	-1.5
	KristenAndSara	-0.6	-0.6	-0.7	-0.6	-0.8	-0.3	-0.5	-0.6	-1.2
	Overall	**-1.5**	**-1.5**	**-1.6**	**-1.3**	**-1.1**	**-1.4**	**-2.5**	**-2.3**	**-2.5**
	Enc. Time[%]	98%			97%			103%		
	Dec. Time[%]	96%			94%			101%		

To verify the performance of the proposed AMVR scheme, we conduct simulations with the common test conditions during HEVC development. The experimental results are shown in Table 1. From Table 1, it can be seen that the average BD-rate gains are 1.5%, 1.3% and 2.5% for luma at RA, LDB and LDP configuration, respectively. Specially, for sequences *BQSquare* and *PartyScene*, the BD-rate gains can achieve up to 17.3% and 7.2% respectively. The reason is that these sequences have rich texture and move slowly, where eighth-pixel resolution is particularly preferred.

For the computation complexity, both the encoding time and decoding time have been reduced to some extent. The time saving mainly comes from the frames where integer or half pixel MV resolution is selected. For these frames, the interpolation and search of quarter subpel can be skipped. On the other hand, the computation of the decision process is simplified.

We also compare the proposed scheme with the latest method used in HM-KTA2.0. Two more cases are tested: without eighth-pixel or without integer-pixel resolution. The total four cases are as follows:

- Case 1: adaptive MV resolution selection from integer to eighth-pixel;
- Case 2: adaptive MV resolution selection from integer to quarter-pixel;
- Case 3: adaptive MV resolution selection from half to eighth-pixel;
- Case 4: adaptive MV resolution on CU level with explicit signal [15].

The results (luma components) are shown in Table 2.

Table 2: The performance comparison with HM16.2 for four test cases.

Class	Random Access (RA)				Lowdelay-B (LDB)				Lowdelay-P (LDP)			
	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 1**	**Case 2**	**Case 3**	**Case 4**	**Case 1**	**Case 2**	**Case 3**	**Case 4**
A	-1.0	-0.8	-0.9	-0.8	-1.0	-0.7	-0.9	-0.7	-0.9	-0.9	-0.9	-0.6
B	-1.2	-1.0	-0.9	-1.0	-1.0	-0.6	-0.8	-0.8	-1.3	-1.0	-1.2	-1.2
C	-1.7	-1.3	-1.7	-0.8	-1.6	-1.3	-1.5	-0.8	-3.1	-1.2	-3.0	-0.6
D	-2.5	-1.6	-2.5	-0.4	-2.2	-1.4	-2.2	-0.5	-5.3	-1.1	-5.3	-0.4
E	-0.8	-0.7	-0.7	-0.7	-0.8	-0.7	-0.6	-0.6	-0.9	-0.9	-0.8	-0.6
Overall	**-1.5**	**-1.1**	**-1.3**	**-0.7**	**-1.3**	**-0.9**	**-1.2**	**-0.7**	**-2.5**	**-1.0**	**-2.1**	**-0.7**

From Table 2, it can be seen that the proposed scheme is preferable than the state-of-the-art method, where the flag on CU level becomes the bottleneck of the performance. Comparing case 1 and case 2, we can find that the eighth-pixel resolution mainly performs well on class C and class D, where some sequences are with rich texture and little motion. The comparison between case 1 and case 3 shows that the integer resolution is seldom to be selected on frame level.

6. Conclusion

In this paper, we present a scheme to select the MV resolution for each frame adaptively. To achieve this goal, we first derive a prediction distortion model in terms of MV resolution, and then a rate model of MVD is developed. To avoid multi-pass encoding, the previous frame in the same temporal layer is used to estimate the rate-distortion cost for each MV resolution. The MV resolution that minimizes the rate-distortion cost on frame level will be selected. Simulation results show that our proposed method works well and can achieve 1.5%, 1.3% and 2.5% BD-rate gains on average for RA, LDB and LDP configuration.

Acknowledgment

This work was supported in part by the Major State Basic Research Development Program of China (2015CB351800), in part by the National Science Foundation (No. 61322106, No. 61571017 and 61272255), and in part by the Scientific Research Common Program of Beijing Municipal Commission of Education (KM201310025009) and Shenzhen Peacock Plan, which are gratefully acknowledged.

References

[1] Marpe D, Wiegand T, Sullivan G J, "The H. 264/MPEG4 advanced video coding standard and its applications," *Communications Magazine, IEEE,* 2006, 44(8): 134-143.

[2] Bross, B., et al, "High efficiency video coding (HEVC) text specification draft 10 (JCTVCL1003)," JCT-VC Meeting (Joint Collaborative Team of ISO/IEC MPEG & ITU-T VCEG). 2013.

[3] McCann K, Han W J, Kim I K, et al, *Samsung's response to the call for proposals on video compression technology*. Joint Collaborative Team on Video Coding (JCT-VC), JCTVC-A124, 2010.

[4] Girod, B, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Transactions on Communications*, vol. 41, no. 4, pp. 604-612, 1993.
[5] Ribas-Corbera J, Neuhoff D L, "Optimizing motion-vector accuracy in block-based video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 4, pp. 497-511, 2001.
[6] Ribas-Corbera, J, Neuhoff D L, "Optimizing block size in motion-compensated video coding," *Journal of Electronic Imagine*, vol. 7, no. 1, pp. 155-165, 1998.
[7] Zhang Q, Dai Y, Ma S, et al, *Rate-Distortion (RD) Analysis of Subpel Motion Vector Resolution Selection for Video Coding*. Conn.: IEEE International Conference on Multimedia and Expo, 2007
[8] Zhang Q, Dai Y, Kuo CCJ, "Direct Techniques for Optimal Sub-Pixel Motion Accuracy Estimation and Position Prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 12, pp. 1735-1744, 2010.
[9] Ji X, Zhao D, Gao W, "Block-Wise Adaptive Motion Accuracy Based B-Picture Coding With Low-Complexity Motion Compensation", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 8, pp. 1085-1090, 2007.
[10] Ma J, An J, Zhang K, Ma S, *Progressive motion vector resolution for HEVC*. Conn.: Visual Communications and Image Processing, 2013.
[11] Wang Z, Ma J, Luo F, Ma S, *Adaptive motion vector resolution prediction in block-based video coding*. Conn.: Visual Communications and Image Processing, 2013.
[12] Chen J, Chen Y, Karczewicz M, et al, *Coding tools investigation for next generation video coding based on HEVC*. Conn.: SPIE Optical Engineering+ Applications. International Society for Optics and Photonics, 2015: 95991B-95991B-9.
[13] Marpe D, Schwarz H, Wiegand T, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 620-636.
[14] Shannon C E. "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, 2001, 5(1): 3-55.
[15] Sullivan GJ, Wiegand T, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 74-90, 1998.
[16] Ma S, Gao W, Lu Y, "Rate-distortion analysis for H. 264/AVC video coding and its application to rate control," *IEEE Transactions on Circuits and Systems for Video Technology*, 2005, 15(12): 1533-1544.