Adaptive Motion Vector Resolution Prediction in Block-Based Video Coding

Zhao Wang[1], Juncheng Ma[1], Falei Luo[1], Siwei Ma[1,2]

[1]*Institute of Digital Media & Cooperative Medianet Innovation Center，Peking University, Beijing, China*

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

{zhaowang, jcma, swma}@pku.edu.cn; falei.luo@vipl.ict.ac.cn

Abstract— **In the classical block-based video coding, motion vector is derived for each coding block to remove the inter-frame redundancy. However, the motion vector resolution is usually restricted to be identical, typically 1/4-pixel resolution, regardless of the different video contents. In this paper, we propose an algorithm that can adaptively select the optimal motion vector resolution at frame level according to the characteristics of the video contents. We first derived a residual energy model, and the major factors that may impact the motion vector resolution are considered, including the texture complexity, motion scale, inter-frame noise and quantization parameter. Experimental results have shown that the proposed scheme can achieve 1.8% BD-rate gain on average without complexity increment.**

Index Terms— **adaptive motion vector resolution, motion vector, motion compensation, rate model, video coding**

I. INTRODUCTION

Block-based motion compensation video coders are widely used because of the excellent performance and reasonable complexity. For each block in the current frame, a similar block searched from the reference frames is served as a prediction block to remove the inter-frame redundancy. The residual, namely the difference between the current block and the prediction block, and the motion vectors denoting the relative position between them are encoded.

The earliest standard H.261 uses integer-pixel motion vector resolution where only prediction at full pixel position can be obtained. However, the motion between consecutive frames is not necessarily integer pixel. Therefore, half-pixel motion vector resolution has been introduced into MPEG-2 and H.263, which significantly improves the coding efficiency. Later, quarter-pixel motion vector resolution was adopted in H.264/AVC, and it is also adopted in the latest H.265/HEVC [1]. While motion vectors with high resolution can provide better prediction, they require more bits to be encoded. The trade-off between the prediction accuracy and the motion vector coding bits needs to be further explored.

Adaptive motion vector resolution has been an active research topic in the past decade. An earliest optimizing motion vector resolution scheme has been proposed by J.Ribas-Corbera in [2] and [3] for H.26L encoders. The author developed a rate model for the motion vector and the residual coding rate to select the optimal motion vector resolution, which minimizes the total bit rates. Nevertheless, this rate model is based on the invariant block size. In the recent work [4], a H.264-based scheme has been proposed. A scheme without overhead bits was presented in [5] for B frame coding, where for all bi-predicted blocks the motion vector resolution is reduced to half-pixel. In [6], a rate-distortion model for motion vector was proposed for power-efficient video coding. In [7] and [8], a progressive motion vector resolution method was proposed, where higher motion vector resolution is employed for motion vectors near to the motion vector predictor (MVP) and lower resolution is employed for motion vectors far from the MVP.

In this paper, we propose a novel method to predict the optimal motion vector resolution for each frame adaptively. To achieve this goal, we first derive an energy model of the prediction residual and then investigate the relationship between the motion vector resolution and the characteristics of the video contents. Based on this study, an approximate model is utilized to determine the optimal motion vector resolution for each frame.

The remaining of this paper is organized as follows: Section II presents the analytical model of the prediction residual. In Section III, the adaptive motion vector resolution selection scheme is proposed. Simulation results are shown in Section IV and Section V concludes this work.

II. AN ANALYTICAL RESIDUAL MODEL

In video coding, motion compensation is performed between the current frame and its reference frames to find the best temporal match. Let $F_t(*)$ and $F_{t-1}(*)$ denote the current frame at instance t and the decoded reference frame at instance t-1, respectively. Then, $F_t(\boldsymbol{s})$ denotes the pixel located at coordinate (s_x, s_y) and $F_{t-1}(\boldsymbol{s}+\boldsymbol{u})$ denotes its corresponding prediction pixel where $\boldsymbol{u}$ represents the ideal motion vector (u_x, u_y) that can be infinite resolution. Hence, we know that

$$F_t(\boldsymbol{s}) = F_{t-1}(\boldsymbol{s}+\boldsymbol{u}) + N(\boldsymbol{s}), \qquad (1)$$

where $N(\boldsymbol{s})$ is interpreted as the inter-frame noise produced by the light change, camera noise, non-translational motion, coding distortion, occlusions, etc. Without such inter-frame noise, the current block and its prediction would be identical.

In practice, we can only use limited motion vector resolution, e.g. 1/2-pixel resolution, 1/4-pixel resolution, or 1/8-pixel resolution, etc. So there exists little gap between the ideal motion vector with infinite resolution and the practical motion vector with limited resolution. If we let $\boldsymbol{u}^*$denote the practical motion vector we used, the prediction of the current pixel $F_t(\boldsymbol{s})$ will be $F_{t-1}(\boldsymbol{s} + \boldsymbol{u}^*)$, not the ideal $F_{t-1}(\boldsymbol{s}+\boldsymbol{u})$. The prediction residual $R(\boldsymbol{s})$ is

$$R(\boldsymbol{s}) = F_{t-1}(\boldsymbol{s} + \boldsymbol{u}^*) - F_t(\boldsymbol{s}). \quad (2)$$

Substituting (1) into (2) yields

$$R(\boldsymbol{s}) = F_{t-1}(\boldsymbol{s} + \boldsymbol{u}^*) - F_{t-1}(\boldsymbol{s}+\boldsymbol{u}) - N(\boldsymbol{s}). \quad (3)$$

Hence, the energy of the residual on a block $\boldsymbol{B}$ is

$$E_{\boldsymbol{B}} = \iint_{\boldsymbol{B}} \left(F_{t-1}(\boldsymbol{s}+\boldsymbol{u}^*) - F_{t-1}(\boldsymbol{s}+\boldsymbol{u}) - N(\boldsymbol{s})\right)^2 dxdy. \quad (4)$$

Expansion the formula as

$$\begin{aligned} E_{\boldsymbol{B}} = &\iint_{\boldsymbol{B}} \left(F_{t-1}(\boldsymbol{s}+\boldsymbol{u}^*) - F_{t-1}(\boldsymbol{s}+\boldsymbol{u})\right)^2 dxdy \\ &-2\iint_{\boldsymbol{B}} \left(F_{t-1}(\boldsymbol{s}+\boldsymbol{u}^*) - F_{t-1}(\boldsymbol{s}+\boldsymbol{u})\right) N(\boldsymbol{s}) dxdy \\ &+ \iint_{\boldsymbol{B}} N(\boldsymbol{s})^2 dxdy. \end{aligned} \quad (5)$$

Considering both $F_{t-1}(\boldsymbol{s}+\boldsymbol{u}^*) - F_{t-1}(\boldsymbol{s}+\boldsymbol{u})$ and noise $N(\boldsymbol{s})$ follow Gaussian distribution, it is anticipated that the second term in (5) will dominate. Hence, the residual energy is

$$\begin{aligned} E_{\boldsymbol{B}} = &\iint_{\boldsymbol{B}} \left(F_{t-1}(\boldsymbol{s}+\boldsymbol{u}^*) - F_{t-1}(\boldsymbol{s}+\boldsymbol{u})\right)^2 dxdy \\ &+ \iint_{B} N(\boldsymbol{s})^2 dxdy. \end{aligned} \quad (6)$$

Let Δ denote the difference between the ideal motion vector $\boldsymbol{u}$ and the practical motion vector $\boldsymbol{u}^*$, namely$\Delta = \boldsymbol{u}^* - \boldsymbol{u}$, or written $(\Delta_x, \Delta_y) = (u_x^* - u_x,\, u_y^* - u_y)$ Based on Taylor series we know that

$$F(\boldsymbol{s}+\Delta) - F(\boldsymbol{s}) = F_x'(\boldsymbol{s})\Delta_x + F_y'(\boldsymbol{s})\Delta_y + o. \quad (7)$$

Using the above Taylor series, the first integral term in (6) is

$$\begin{aligned} &\iint_{\boldsymbol{B}} \left(F_{t-1}(\boldsymbol{s}+\boldsymbol{u}^*) - F_{t-1}(\boldsymbol{s}+\boldsymbol{u})\right)^2 dxdy \\ &= \iint_{\boldsymbol{B}} \left(a\Delta_x^2 + b\Delta_y^2 + 2c\Delta_x\Delta_y\right) dxdy, \end{aligned} \quad (8)$$

where a, b, c depend on the pixel gradient

$$a = (\frac{\partial}{\partial x})^2 \quad b = (\frac{\partial}{\partial y})^2 \quad c = \frac{\partial}{\partial x}\frac{\partial}{\partial y}.$$

Assuming Δ_x and Δ_y follow Gaussian distribution, the third integral term in (8) will also be zero. For this reason, formula (8) will be

$$\begin{aligned} &\iint_{\boldsymbol{B}} \left(F_{t-1}(\boldsymbol{s}+\boldsymbol{u}^*) - F_{t-1}(\boldsymbol{s}+\boldsymbol{u})\right)^2 dxdy \\ &= \iint_{\boldsymbol{B}} \left(a\Delta_x^2 + b\Delta_y^2\right) dxdy. \end{aligned} \quad (9)$$

Substituting (9) into (6), the residual energy of one block is,

$$E_{\boldsymbol{B}} = \iint_{\boldsymbol{B}} \left(a\Delta_x^2 + b\Delta_y^2\right) dxdy + \iint_{\boldsymbol{B}} N(\boldsymbol{s})^2 dxdy. \quad (10)$$

Considering Δ_x and Δ_y are related with the motion vector resolution we used, Equation (10) can be approximately transform into

$$E_B = A_B\Delta^2 + N_B, \quad (11)$$

where $\boldsymbol{B}$ represents one coding block. $E_{\boldsymbol{B}}$ is the energy of the prediction residual, and $A_{\boldsymbol{B}}, N_{\boldsymbol{B}}$ represent the pixel gradient and inter-frame noise, respectively.

From the above model, we know that the energy of the prediction residual is decided by the pixel gradient, motion vector resolution and the inter-frame noise. Considering the image texture complexity reflects the pixel gradient, we infer that the optimal motion vector resolution of one coding frame is related with its texture complexity and inter-frame noise.

III. Adaptive motion vector resolution scheme

From the residual model presented at Section II, we know that texture complexity and inter-frame noise impact on the optimal motion vector resolution. Nevertheless, there exist other influencing factors, because not only residuals but also the motion vectors need to be encoded. In this section, we search more factors and explore the numerical relationship between them and the selection of motion vector resolution.

A. The Influencing Factors on Optimizing Motion Vector Resolution

In the current H.265/HEVC codec, rate-distortion optimization (RDO) is performed by minimizing the following Lagrangian cost function [9]

$$J(\boldsymbol{s}, \boldsymbol{c} \mid \lambda) = D(\boldsymbol{s}, \boldsymbol{c}) + \lambda R_{\text{mv}}(\boldsymbol{s}, \boldsymbol{c}), \quad (12)$$

where $\boldsymbol{s}$ and $\boldsymbol{c}$ are the original and corresponding prediction block, $D(*)$ reflects the residual and $R_{\text{mv}}(*)$ means the bits of encoding the motion vectors. λ is the Lagrangian multiplier which is related to the quantization parameter (*QP*).

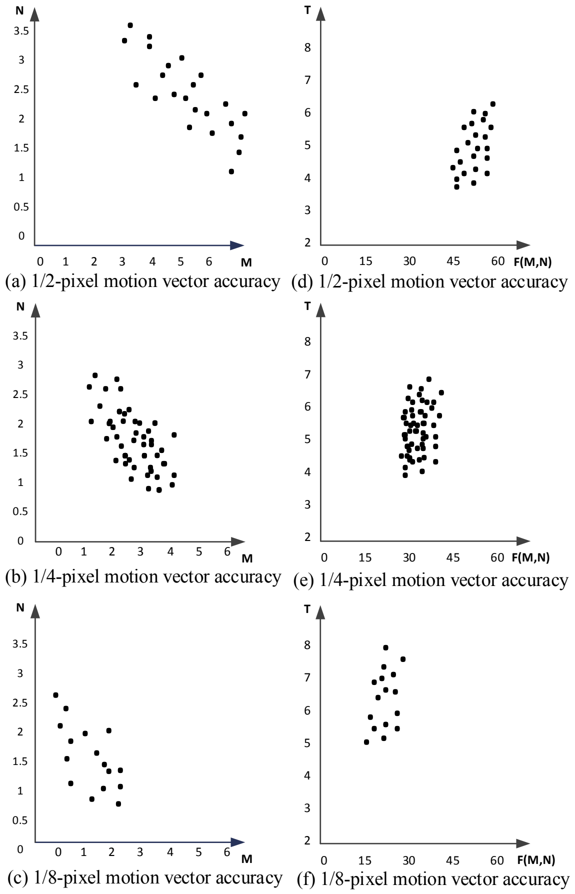
Generally, large motion vector need more bits to be encoded and small motion vector need less bits. With low *QP*, residuals take majority of the bit rate. On the contrary, motion vector coding bits occupy more with high *QP*. Combining our residual model with the RDO function, we observe that the key factors which impact on optimizing motion vector resolution are texture complexity, inter-frame noise, motion scale and *QP*.

B. Impact of Texture Complexity, Inter-frame Noise and Motion Scale with Fixed QP

To explore the mathematical relationship between these factors and the optimal motion vector resolution, we first maintain fixed *QP* because it is not related with the video contents. We calculate the texture complexity (denoted as *T*), inter-frame noise (denoted as *N*) and motion scale (denoted as *M*) using the following formulas respectively

$$\begin{aligned} T &= \frac{1}{2n}\sum\left(\left(s_{x,y} - s_{x+1,y}\right)^2 + \left(s_{x,y} - s_{x,y+1}\right)^2\right), \\ N &= \frac{1}{n}\sum\left|s_{x,y} - s'_{x+u,y+v}\right|, \\ M &= \frac{1}{n}\sum\left(|mvd_x| + |mvd_y|\right) * pu_{size}, \end{aligned} \quad (13)$$

where n is the number of total pixels in one frame. $s_{x,y}$ represents each pixel in the current frame. $s_{x+1,y}$, $s_{x,y+1}$, $s'_{x+u,y+v}$ are the right pixel, below pixel and prediction pixel, respectively. mvd_x and mvd_y represent the *x* and *y*



(a) 1/2-pixel motion vector accuracy (d) 1/2-pixel motion vector accuracy

(b) 1/4-pixel motion vector accuracy (e) 1/4-pixel motion vector accuracy

(c) 1/8-pixel motion vector accuracy (f) 1/8-pixel motion vector accuracy

Fig. 1 The distribution of the value of motion scale (*M*), inter-frame noise (*N*), and texture complexity (*T*) with different optimal motion vector resolution

components of motion vectors in each block whose size is denoted as pu_size.

We test 100 video sequences (not containing HEVC test sequences) with motion vector encoded by 1/2-pixel, 1/4-pixel and 1/8-pixel resolution at QP 32, respectively, and calculate the value of texture complexity, inter-frame noise and motion scale for each sequence by formula (13). Observing the test results shown in Fig. 1, we find that the values of these three factors locate at different area among sequences with different optimal motion vector resolution, and converge near among sequences with the same optimal motion vector resolution.

Fig.1 (a) shows that most of the sequences whose optimal motion vector resolution is 1/2-pixel have large motion and inter-frame noise. On the contrary, the sequences with 1/8-pixel as the optimal motion vector resolution shown in Fig.1 (c) have little motion and inter-frame noise. Other sequences with moderate motion and noise perform best with 1/4-pixel motion vector resolution.

According to the left column in Fig.1, the impact of motion scale and inter-frame noise on optimizing motion vector resolution can be expressed mathematically as

$$F(M,N) = 12M + 7N. \quad (14)$$

The simulations results of $F(M,N)$ are shown in the right column in Fig.1. These charts show that the motion vector resolution need increase where more texture is present and decrease when there is much inter-frame noise or large motion. The statistics confirm that motion scale, inter-frame noise and texture complexity directly impact on the optimal motion vector resolution. The weight of each factor can be modelled by

$$F(M,N,T) = 12M + 7N - 5T. \quad (15)$$

C. Impact of QP on Optimizing Motion Vector Resolution

The optimal motion vector resolution of one sequence will vary when encoded with different *QP*. Generally speaking, it tends to select high motion vector resolution at high bit rate and choose low resolution at low bit rate. We model the impact of *QP* as

$$F(M,N,T;QP) = F(M,N,T;(Th_0,Th_1)). \quad (16)$$

where the two thresholds (Th_0,Th_1) just related with *QP* are used to control the selection of motion vector resolution. *F*(*M*, *N*, *T*; (*Th*$_0$, *Th*$_1$)) is defined as follows,

$$F(M,N,T;(Th_0,Th_1)) = \begin{cases} 1/8\, pixel\ resolution & when\ F(M,N,T) \leq Th_0 \\ 1/4\, pixel\ resolution & when\ Th_0 \leq F(M,N,T) \leq Th_1 \\ 1/2 pixel\ resolution & when\ F(M,N,T) \geq Th_1. \end{cases} \quad (17)$$

We test 32 sequences encoded at different *QP* points from 18 to 46 to model the relation between the thresholds (Th_0,Th_1) and *QP*. The test results are depicted in Fig.2

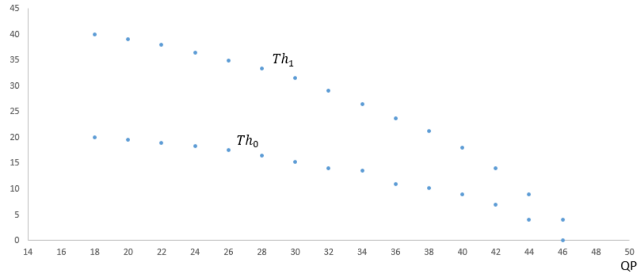


Fig. 2 The statistics of thresholds $(\boldsymbol{Th_0},\boldsymbol{Th_1})$

The statistics of thresholds can be modelled as

$$Th_0 = a_0QP^2 + b_0QP + c_0,$$
$$Th_1 = a_1QP^2 + b_1QP + c_1, \quad (18)$$

where a_0, b_0, and c_0 are approximately -0.02, 0.57, and 16, respectively; a_1,b_1 and c_1 are approximately -0.03, 0.6 and 38 respectively by curve-fitting the optimal thresholds.

D. The Adaptive Motion Vector Resolution Scheme

Based on the above study, we propose an adaptive motion vector resolution scheme. For each frame to be encoded, firstly, obtain QP and calculate its texture complexity (*T*) directly, then calculate the motion scale (*M*) and inter-frame noise (*N*) of the previous encoded frame. Compute the value of $F(M,N,T)$ by (15) and compare it with the thresholds obtained by (18) to select the optimal motion vector resolution. The encoder will use two bits to encode the index of selected motion vector resolution and transfer it to the decoder. The flowchart of proposed scheme is shown as the following Fig. 3.

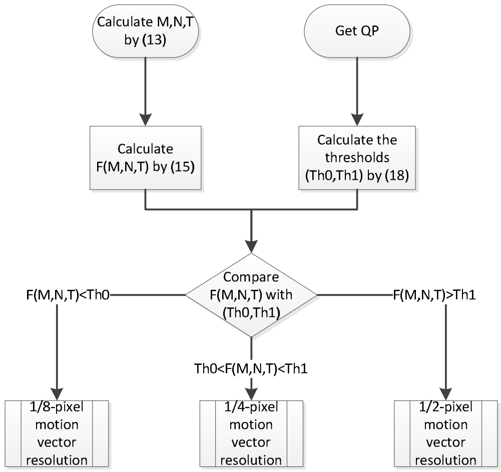

Calculate M,N,T by (13)
Get QP
Calculate F(M,N,T) by (15)
Calculate the thresholds (Th0,Th1) by (18)
Compare F(M,N,T) with (Th0,Th1)
F(M,N,T)<Th0
F(M,N,T)>Th1
Th0<F(M,N,T)<Th1
1/8-pixel motion vector resolution
1/4-pixel motion vector resolution
1/2-pixel motion vector resolution

Fig. 3 The flowchart of the proposed scheme

IV. Experimental results

The proposed adaptive motion vector resolution scheme is integrated into HEVC reference software HM16.2. Simulations are conducted for ten test sequences with various video characteristics. The experimental results are shown in Table I.

TABLE I The Performance of Proposed Scheme

Sequence	Random Access			Lowdelay-B			Lowdelay-P		
	Y	U	V	Y	U	V	Y	U	V
BlowingBubbles	-1.3	-1.6	-1.5	-1.4	-1.5	-1.0	-2.8	-3.2	-3.0
BQSquare	-5.9	-3.7	-5.0	-5.7	-4.4	-6.1	-14.6	-11.3	-13.4
BasketballDrill	-1.8	-2.6	-2.3	-2.0	-4.3	-3.3	-2.9	-5.3	-4.3
BQMall	-1.3	-1.4	-1.3	-1.6	-1.5	-1.1	-2.6	-2.5	-2.2
PartyScene	-2.6	-2.9	-2.7	-2.7	-3.2	-3.4	-6.7	-5.5	-5.9
Fourpeople	-1.2	-0.4	-0.7	-0.6	-0.3	-0.0	-0.5	-0.1	-0.3
PeopleOnStreet	-1.1	-2.1	-1.7	-1.0	-1.8	-1.1	-0.4	-1.5	-0.9
Kimono	-1.3	-1.3	-1.0	-1.1	-0.6	-0.9	-1.1	-1.4	-1.0
BasketballDrive	-0.5	-1.0	-0.9	-0.7	-1.2	-1.0	-0.4	-1.3	-0.6
Cactus	-0.7	-0.5	-0.5	-0.5	-0.3	-0.2	-0.9	-0.7	-0.8
Average	**-1.8**	**-1.8**	**-1.9**	**-1.7**	**-1.9**	**-1.8**	**-3.3**	**-3.2**	**-3.3**
Enc. Time[%]	93%			92%			101%		
Dec. Time[%]	100%			98%			102%		

From Table I, it can be seen that the proposed method can achieve 1.8%, 1.7% and 3.3% BD-rate gain on average for Random Access, Lowdelay-B and Lowdelay-P configuration, respectively. Specifically, the coding gain is much higher for sequences with large motion or rich texture, because most frames perform well using 1/2-pixel or 1/8-pixel motion vector resolution.

To further verify the performance of the proposed method, we build a multi-pass codec which encodes each frame with 1/2-pixel, 1/4-pixel and 1/8-pixel resolution respectively, and then select the best one as the final resolution for this frame. Taking this multi-pass codec as anchor, we estimate the accuracy of proposed method by the percentage of frames whose motion vector resolution are same in anchor and in codec integrated with proposed method. The results are shown in Table II and we can see that our method is reliable for most frames.

TABLE III The Accuracy of Proposed Scheme at RA Configuration

Sequence	Accuracy	Sequence	Accuracy
BlowingBubbles	94.7%	Fourpeople	93.3%
BQSquare	97.1%	PeopleOnStreet	89.8%
BasketballDrill	88.4%	Kimono	92.0%
BQMall	93.6%	BasketballDrive	91.3%
PartyScene	95.3%	Cactus	90.2%

As for the complexity, the proposed scheme introduces negligible computation for predicting the optimal motion vector resolution. Even time saving has been obtained for the encoder because of the selection of 1/2-pixel resolution in partial frames.

V. Conclusions

In this paper, we have derived an energy model of the prediction residual. The key factors impacting on the optimal motion vector resolution are analyzed, including the texture complexity, motion scale, inter-frame noise and quantization parameter. Moreover, an adaptive motion vector resolution scheme has been proposed and integrated into HEVC codec. Experimental results show that our method can achieve 1.8~3.3% BD-rate gain on average.

Acknowledgment

This work was supported in part by the National High-tech R&D Program of China (863 Program, 2015AA015903), National Natural Science Foundation of China (61322106, 61210005), and Shenzhen Peacock Plan, which are gratefully acknowledged.

References

[1] B. Bross, et al., “High Efficiency Video Coding (HEVC) text specification draft 10 (for FDIS & Consent)” ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) document, JCTVC-L1003, Jan. 2013.

[2] Ribas-Corbera J, Neuhoff D L. “Optimizing motion-vector accuracy in block-based video coding,”[J]. *Circuits and Systems for Video Technology, IEEE Transactions on*, 2001, 11(4): 497-511.

[3] Ribas-Corbera J, Neuhoff D L. “Optimizing block size in motion-compensated video coding,”[J]. *Journal of Electronic Imaging*, 1998, 7(1): 155-165.

[4] Corrado S, Agostini M A, Cagnazzo M, et al. “Improving H. 264 performances by quantization of motion vectors,”[C]//Picture Coding Symposium, 2009. PCS 2009. IEEE, 2009: 1-4.

[5] Ji X, Zhao D, Gao W. “Block-wise adaptive motion accuracy based b-picture coding with low-complexity motion compensation,”[J]. *Circuits and Systems for Video Technology, IEEE Transactions on*, 2007, 17(8): 1085-1090.

[6] Zhang Q, Dai Y, Ma S, et al. “Rate-Distortion (RD) Analysis of Subpel Motion Vector Resolution Selection for Video Coding,” [C]//Multimedia and Expo, 2007 IEEE International Conference on. IEEE, 2007: 380-383.

[7] Karczewicz M, Chen P, Joshi R, et al. “Video coding technology proposal by Qualcomm Inc,”[J]. JCTVC Contribution JCTVC-A121, Dresden, Germany, 2010.

[8] Ma J, An J, Zhang K, et al. “Progressive motion vector resolution for HEVC,”[C]//Visual Communications and Image Processing (VCIP), 2013. IEEE, 2013: 1-6.

[9] Sullivan G J, Wiegand T. “Rate-distortion optimization for video compression,”[J]. *Signal Processing Magazine*, IEEE, 1998, 15(6): 74-9