Second Order Residual Prediction for HEVC Inter Coding

Chia-Hung Yeh, Cheng-Wei Lee, Shu-Jhen Fan Jiang, Yu-Hsien Sung, and Wen-Jung Huang
Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan, R.O.C.
E-mail: yeh@mail.ee.nsysu.edu.tw Tel: +886-7-5252000 #4112

***Abstract*—In previous researches, the reconstructed residual data in the current frame or in the reference frame are used to predict the residual of the current block. These researches demonstrate that the redundancy still exists in the residual. We propose a new residual prediction scheme for inter coding to further remove the redundancy in the residual encoded by the original framework so that the coding performance can be enhanced. Experimental results show the proposed method achieves an improvement of 4.26% in bitrate saving; meanwhile, the encoding time of proposed method is only increased by 7.95% when compared to HEVC (High Efficiency Video Coding).**

I. INTRODUCTION

H.264/AVC was the most powerful open video coding standard with an increase of 50% in coding performance when compared to MPEG-2. HEVC, the latest video coding, completed in 2013, provides 50% bitrate reduction when compared to H.264/AVC. The invention of new and innovative technologies enables the progress in the development of the next-generation video coding standard. It takes seven years from the finalization of H.264/AVC to the start of the HEVC standard activity, reflecting the difficulty of video coding advancement. To improve the coding performance, many schemes including coding tree unit, advanced motion vector prediction, and motion merging are applied to HEVC. In HEVC, more neighboring blocks' MVs are used to predict that of the current block. The CU (Coding Unit) of HEVC varies from depth 0 (64×64) to depth 3 (8×8). Then each CU is split into one to four PUs, and the motion estimation is performed in each PU. The predicted MV of the current PU is the medium value of the MVs of neighboring blocks in the current frame and the MV of the co-located block in the reference frame. To avoid transmitting MVs, the MV of the current PU can be merged with that of the adjacent PU.

Instead of using motion compensation in inter frame, Suzuki *et al.* [1] and Chen *et al.* [2] proposed a patch matching for inter frame prediction based on template matching prediction to improve the efficiency at both of low and high bitrates. These kinds of template matching algorithms use the coded pixels in the "L"-shape in the current frame as the search pattern to find candidates in the reference frame which are similar to the coded pixels in the current frame. Then, these candidates are synthesized to form one prediction block. Some researches proposed new transforms to efficiently compact the energy of residual and improving the coding efficiency of video coders [3]-[5]. Zhang *et al.* [3] proposed a spatially varying transform which varies the position of the transform block and the transform size. Zhao *et al.* [4] and Xu *et al.* [5] used the concept of KLT (Karhunen–Loève Transform) to compact energy of the residual. Zhao *et al.* [4] proposed a rate-distortion optimized transform which provides multiple transform basis functions. The residual is transformed by the best set among the transform basis functions. Xu *et al.* [5] proposed a feature matching algorithm to find the best transform basis. Residual prediction is another way to improve the coding efficiency. Kin *et al.* [6] used the residual of adjacent coded blocks as the reference of the second prediction. Instead of using adjacent blocks as the reference, Zhang *et al.* [7] proposed a MOR (Multi-Order-Residual) coding approach in the frequency domain for high bitrate video compression.

The rest of the paper is organized as follows. Section II analyses the relationship between the bitrate and residual with different quantization parameters (QPs) for HEVC. The proposed second order residual prediction is explained in Sec. III. Section IV shows the experimental results of the proposed algorithm. Finally, concluding remarks are given in Sec. V.

II. ANALYSIS OF BITRATE AND RESIDUAL FOR DIFFERENT QPS IN HEVC

This subsection analyzes the relationship between bitrate and residual with different QPs (Quantization Parameters) in HEVC. The testing sequences for all experiments include *ParkScene* (1920×1080), *BasketballDrill* (832×480), *RaceHorses* (416×40), and *vidyo1* (1280×720) and each sequence encodes 100 frames. The QPs are set from 22 to 37 in intervals of 5.

The relationship between the bitrate and the QP is shown in Fig. 1. According to our observation, the coding performance of HEVC is very effective when the QP is coarse while the bitrate increases dramatically when the QP gets finer. Therefore, the coding performance of HEVC in high bitrate can be further improved. To further understand the relationships between residual bits and header bit, several testing sequences are analyzed and the results are shown in Fig. 2 The residual bits include CBP (Coded Block Pattern) information and the coefficients of luma and chroma. The header bits include the MVs, CU split information, PU partitions, and the merge mode information. In Fig. 2, the growth of header bits is slower than that of residual bits when the QP is finer. Figure 3 shows that the required bits for encoding residual are around 50% of the total bits when QP is 37, and up to 80% when QP is 22. Our observations show HEVC provides excellent rate performance at low to medium bitrate coding and there is a room for improvement at high bitrate. Hence, we propose a second order residual prediction algorithm in the pixel domain to improve the coding performance of HEVC.

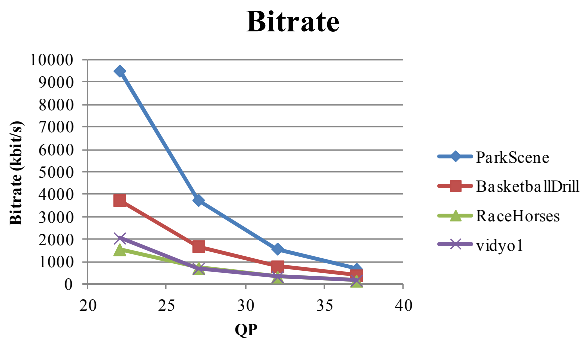

Fig. 1 Illustration the relationship between the bitrate and QP

III. Proposed Second Order Residual Prediction

Based on our observations, a second order residual prediction is proposed to reduce the bitrate. In our assumption, the temporal redundancy of residual exists between the current frame and its reference frame. Therefore, a second order residual prediction algorithm is proposed for bitrate reduction in HEVC. Figure 4 illustrates the architecture of the proposed algorithm. First, motion estimation in each PU partition is performed in the current CU to find the first order residual. Then, the proposed algorithm is performed to find the second order residual. A rate-distortion cost function is used to check if the proposed method has better coding performance than the original HEVC inter coding in the current CU. An additional MV is required to transmit to the decoder to indicate the position of the predicted residual block.

The first order residual predicts the difference between the reference frame and the current frame. The first order residual block, $r_t^{(1)}$ in the t^{th} frame, is defined as Eq. (1)

$$r_t^{(1)}(i,j) = I_t(i,j) - \tilde{I}_{t-n}(i+v_x^{(1)}, j+v_y^{(1)}), \quad (1)$$

where $(v_x^{(1)}, v_y^{(1)})$ is the MV of the first order residual, (i,j) is the position, I_t is the current frame and $\tilde{I}_{t-n}$ is the reconstructed reference frame. After the first order residual prediction, the second order residual block is acquired by finding the difference between the first order residual block and the reconstructed residual in the reference frame. Similar to the first order residual, the second order residual block, $r_t^{(2)}$, can be defined as Eq. (2).

$$r_t^{(2)}(i,j) = r_t^{(1)}(i,j) - \tilde{r}_{t-n}(i+v_x^{(2)}, j+v_y^{(2)}), \quad (2)$$

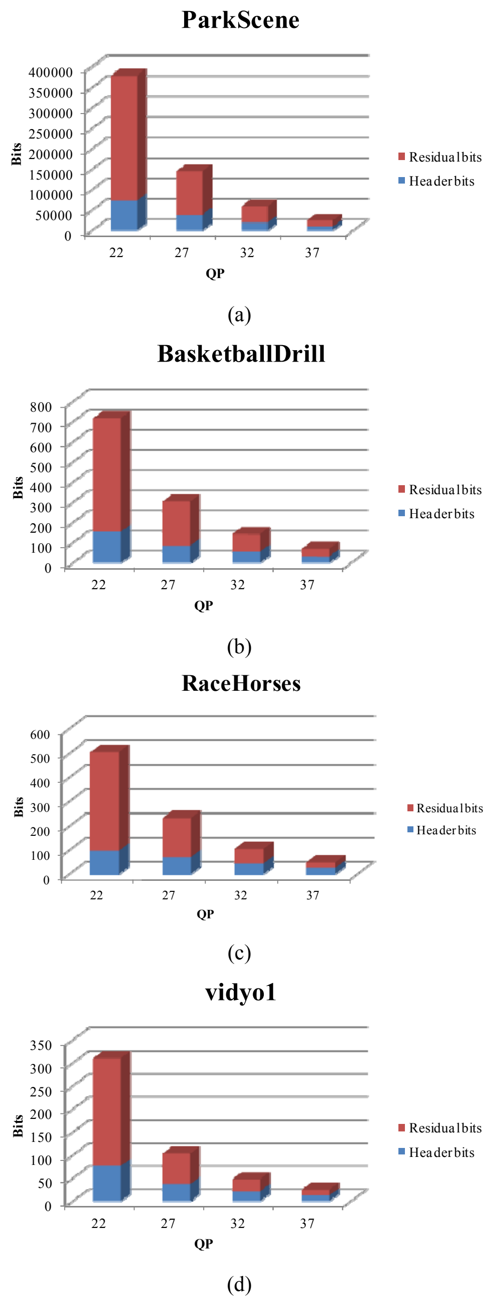

Fig. 2 Header bits and the residual bits in (a) *ParkScene* (b) *BasketballDrill* (c) *RaceHorses* and (d) *vidyo1* sequence

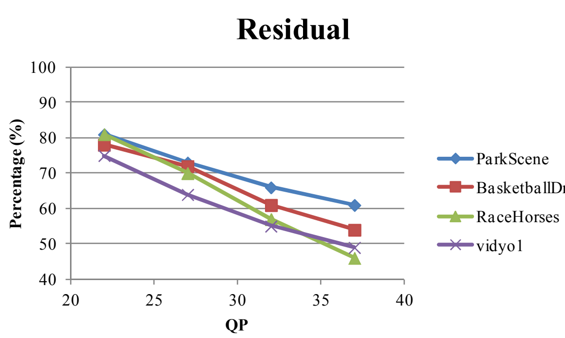

Fig. 3 Ratio of residual bits to total bits

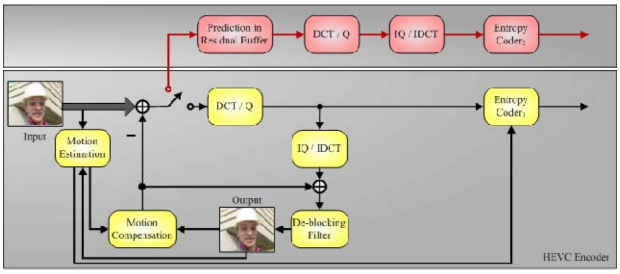

Fig. 4 Architecture of the proposed second order residual prediction algorithm

where $\tilde{r}_{t-n}$ is the reconstructed residual frame in the reference frame and $(v_x^{(2)}, v_y^{(2)})$ is the MV of the second order residual. To find the tradeoff between the prediction error and the required bits of encoding MVs of the second order residual, the Lagrangian cost function for motion estimation J_{motion} is applied.

$$\begin{cases} J_{motion}\left(r_t^{(2)},\overrightarrow{v^{(2)}},\lambda_{motion}\right) = SAD\left(r_t^{(2)}\right) + \lambda_{motion} \times R\left(\overrightarrow{v_d^{(2)}}\right) \\ SAD\left(r_t^{(2)}\right) = \sum_{m=0}^{M}\sum_{n=0}^{N}\left|r_t^{(2)}(i+m,j+n)\right| \\ \overrightarrow{v_d^{(2)}} = \overrightarrow{v^{(2)}} - \overrightarrow{v_p^{(2)}} \end{cases}, \quad (3)$$

where $\overrightarrow{v^{(2)}}$ represents the MV of the second order residual, λ_{motion} is the Lagrangian multiplier, SAD is the sum of the absolute difference of the second order residual, R is the constrained condition, $\overrightarrow{v_d^{(2)}}$ is the MV difference of the second order residual, and $\overrightarrow{v_p^{(2)}}$ is the predicted MV of the second order residual. The predicted MV of the second order residual is set as the best matching MV of the first order residual. After the motion estimation and motion compensation in the second order residual block, transform coding and inverse transform coding are employed to find the reconstructed second order residual block, $\tilde{r}_t^{(2)}$. The Lagrangian cost function for the residual coding is used to determine the better coding performance from predictions with and without residual predictions as shown in Eqs. 4 and 5, respectively.

$$J_{mode}^{(1)}\left(\tilde{r}_t^{(1)},\overrightarrow{v^{(1)}},\lambda_{mode}\right) = SSE\left(\tilde{r}_t^{(1)},\overrightarrow{v^{(1)}}\right) + \lambda_{mode} \times R\left(\overrightarrow{v_d^{(1)}}\right), \quad (4)$$

$$J_{mode}^{(2)}\left(\tilde{r}_t^{(2)},\overrightarrow{v^{(1)}},\overrightarrow{v^{(2)}},\lambda_{mode}\right) = SSE\left(\tilde{r}_t^{(2)},\overrightarrow{v^{(1)}},\overrightarrow{v^{(2)}}\right) + \lambda_{mode} \times \left(R\left(\overrightarrow{v_d^{(1)}}\right) + R\left(\overrightarrow{v_d^{(2)}}\right)\right), \quad (5)$$

where $J_{mode}^{(1)}$ represents the mode without residual prediction, $J_{mode}^{(2)}$ represents the mode with residual prediction, $\overrightarrow{v^{(1)}}$ is the MV of first order residual, $\tilde{r}_t^{(1)}$ and $\tilde{r}_t^{(2)}$ represents the reconstructed first order residual block and the reconstructed second order residual block, respectively, SSE is the sum of squared error, and $\overrightarrow{v_d^{(1)}}$ is the MV difference of first order residual. The prediction mode with the smallest cost function is selected as the best mode and the reconstructed residual frame in current frame can be represented as

$$\tilde{r}_t(i,j) = \begin{cases} \tilde{r}_t^{(1)}, & \text{if } J_{mode}^{(1)} \le J_{mode}^{(2)} \\ \tilde{r}_t^{(2)}, & \text{otherwise} \end{cases}. \quad (6)$$

IV. Experimental Results

The proposed method is implemented in HEVC reference software HM10.0. Eighteen testing sequences are encoded with HM10.0 and the proposed method. BDBR (Bjontegaard Delta BitRate) and BDPSNR (Bjontegaard Delta PSNR) [8] are used to evaluate the coding gain of proposed method when compared to HM10.0. The main profile with the low-delay P coding configuration is used in this simulation. The increasing time is used to compare the computational complexities of the proposed method and of HM 10.0. The increasing time is defined as

$$\text{Increasing Time} = \frac{T_{proposed} - T_{HM}}{T_{HM}} \times 100\%, \quad (7)$$

where T_{HM} and $T_{proposed}$ represent the encoding time of HM 10.0 and the proposed method, respectively. Table I shows the proposed method outperforms HEVC by 4.26% savings in bitrate or an increase of 0.14 dB PSNR. The encoding time of the proposed method is only increased by 7.95% when compared to HM10.0. Figures 5-8 show the rate-distortion performance of the proposed methods and HM 10.0 in *PeopleOnStreet*, *BasketballDrill*, *RaceHorses*, and *vidyo1* sequences, respectively. In these figures, the proposed method has better coding performances than HM10.0.

TABLE I
COMPARISONS OF BDBR (%), BDPSNR (dB), AND INCREASING TIME (%)

Sequence	class	BDBR (%)	Proposed BDPSNR (dB)	Increasing Time (%)
Traffic	A	-4.65	0.13	11.22
PeopleOnStreet	A	-10.44	0.42	5.54
Kimono1	B	-2.65	0.08	7.46
ParkScene	B	-4.08	0.11	5.68
Cactus	B	-3.11	0.06	7.39
BasketballDrive	B	-2.91	0.06	6.93
BQTerrace	B	-0.99	0.02	10.90
BasketballDrill	C	-4.94	0.21	9.85
BQMall	C	-4.50	0.16	10.15
PartyScene	C	-2.49	0.09	8.63
RaceHorses	C	-3.72	0.14	9.15
BasketballPass	D	-4.99	0.21	4.19
BQSquare	D	-1.51	0.06	11.42
BlowingBubbles	D	-2.83	0.10	7.13
RaceHorses	D	-5.85	0.25	5.60
vidyo1	E	-6.31	0.17	4.10
vidyo3	E	-4.95	0.15	9.55
vidyo4	E	-5.69	0.14	8.29
average		-4.26	0.14	7.95

V. Conclusions

A second order residual prediction algorithm to improve the efficiency of inter coding is proposed. According to the analyses of relationship between bitrate and residual with different QPs, there is a room for the improvement of HEVC coding performance at high bitrate. Therefore, the proposed algorithm further removes the redundancy of the original residuals via their temporal correlation. Experimental results show that the proposed algorithm reduces 4.26% bitrate when compared to HEVC.

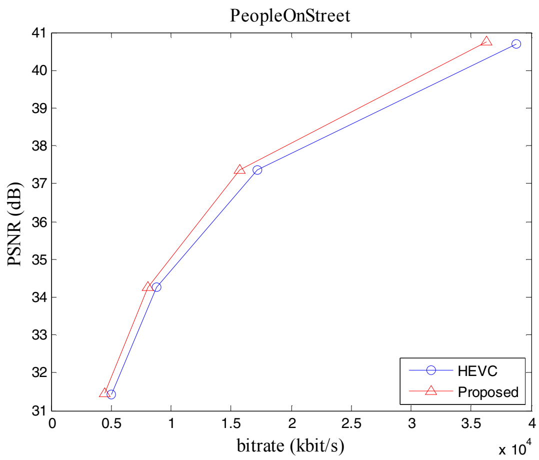

Fig. 5 RD performance of HM10.0 and the proposed method for *PeopleOnStreet*

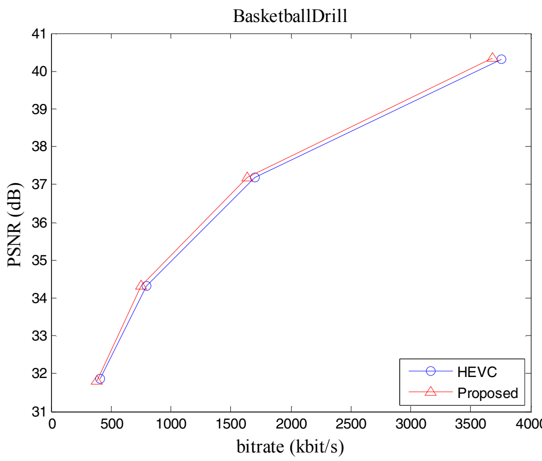

Fig. 6 RD performance of HM10.0 and the proposed method for *BasketballDrill*

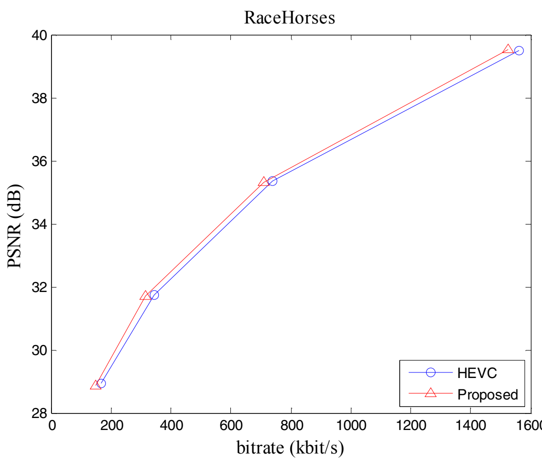

Fig. 7 RD performance of HM10.0 and the proposed method for *RaceHorses*

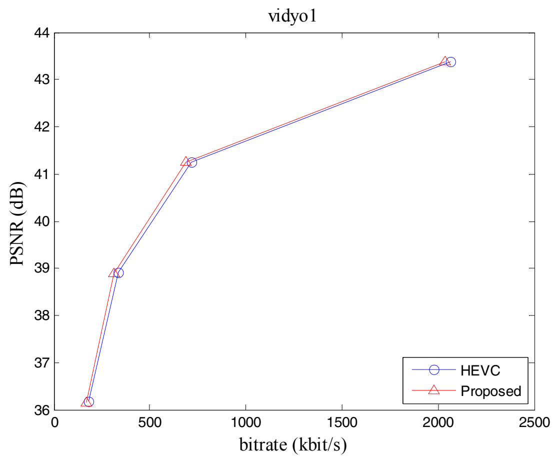

Fig. 8 RD performance of HM10.0 and the proposed method for *vidyo1*

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Science and Technology (MOST) for financially supporting this research under contracts No. MOST102-2221-E-110-032-MY3, MOST103-2221-E-110-045-MY3, and MOST103-3113-E-110-002.

REFERENCES

[1] Y. Suzuki, C. Seng Boon, and T. K. Tan, “Inter frame coding reference with template matching averaging,” in *Proceedings of IEEE International Conference on Image Processing*, pp. III-409-III-412, 2007.

[2] T. Chen, X. Sun, and F. Wu, “Predictive patch matching for inter frame coding,” in *Proceedings of SPIE Visual Communications and Image Processing*, vol. 7744, pp. 774412-1-774412-8, 2010.

[3] C. Zhang, K. Ugur, J. Lainema, A. Hallapuro, and M. Gabbouj, “Video coding using spatially varying transform,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 2, pp. 127-140, Feb. 2011.

[4] X. Zhao, L. Zhang, S. Ma, and W. Gao, “Video coding with rate-distortion optimized transform,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 1, pp. 138-151, 2012.

[5] L. Xu and K. N. Ngan, “Video content dependent directional transform for high performance video coding,” in *Proceedings of IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pp. 79-83, 2012.

[6] S. D. Kim, S. J. Hwang, and M. H. Sunwoo, “Novel residual prediction scheme for hybrid video coding,” in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pp. 617-620, 2009.

[7] Q. Zhang, S.-H. Kim, Y. Dai and C.-C. J. Kuo, “Multi-Order-Residual (MOR) video coding: framework, analysis and performance,” in *Proceedings of SPIE Visual Communications and Image Processing*, vol. 7744, 2010.

[8] G. Bjontegaard, “Calculation of average PSNR differences between RD-curves,” ITU-T Q6/SG16, Doc. VCEG-M33, Apr. 2001.