FAST CODING TREE UNIT DEPTH DECISION FOR HIGH EFFICIENCY VIDEO CODING

Zhaoqing Pan[1,2] *Sam Kwong*[2,3]⋆ *Yun Zhang*[2] *Jianjun Lei*[4] *Hui Yuan*[2]

[1] School of Computer Science and Engineering, Hebei University of Technology, Tianjin, China
[2] Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong
[3] City University of Hong Kong Shenzhen Research Institute, Shenzhen, China
[4] School of Electronic Information Engineering, Tianjin University, Tianjin, China
panzhaoqing@scse.hebut.edu.cn, {cssamk, yunzhang, huiyuan}@cityu.edu.hk, jjleitg@gmail.com

ABSTRACT

High Efficiency Video Coding (HEVC) is the latest video coding standard, which adapts quadtree structure based Coding Tree Unit (CTU) to improve the coding efficiency. In HEVC encoding process, the CTU is recursively partitioned into coding units according to the quadtree depth. This technique increases the coding efficiency of HEVC, however, the achieved coding efficiency comes at the cost of high computational complexity. In this paper, we propose a fast CTU quadtree depth decision algorithm to reduce the computational complexity of HEVC. Firstly, based on the best CTU depth correlation among spatial and temporal neighboring CTUs, an early quadtree depth 0 decision algorithm is proposed. Then, according to the correlation between the prediction unit mode and the best CTU depth selection, a quadtree depth 3 skipped decision algorithm is proposed. Experimental results show that the proposed algorithm can achieve 40% on average encoding time saving, while maintaining a comparable rate-distortion performance.

Index Terms— Coding tree unit, quadtree depth, prediction unit, fast algorithm, HEVC, video coding

1. INTRODUCTION

High Efficiency Video Coding (HEVC) is the latest video coding standard, which is proposed by the Joint Collaborative Team on Video Coding (JCT-VC) under ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG), and is designed to address the issues that the increased video resolution and increased use of parallel processing architectures [1, 2]. HEVC can achieve equivalent perceptual visual quality as H.264/AVC while only using about 50% bit rate [1]. However, the coding gain comes at the cost of these high computational complexity coding tools, including quadtree based Coding Tree Unit (CTU), large and asymmetric Prediction Unit (PU), advanced motion vector prediction, new intra prediction methods and so on. Of all these advanced coding tools, the quadtree based CTU encoding process consumes the largest proportion of the total encoding time. Hence, if the CTU encoding process is simplified, the computational complexity of HEVC will be reduced significantly.

Recently, many researchers have devoted their efforts to reduce the computational complexity of the quadtree structure based CTU in HEVC. Choi *et al.* proposed a simple tree-pruning algorithm that the CTU coding process will be terminated if the PU mode of current Coding Unit (CU) is the SKIP mode [3]. In [4], according to the CTU depth information of previously coded slices and CTUs, Li *et al.* proposed an adaptive CTU depth range algorithm. Based on the average Rate Distortion (RD) cost of previous skipped CUs, Kim *et al.* proposed an adaptive CU early termination algorithm for HEVC [5]. Based on the Bayesian decision rule, Shen *et al.* proposed a fast CU size decision algorithm [6]. In [7], an fast CU size decision method is proposed by using the characteristics of motion homogeneity, RD cost and SKIP mode. Based on the learning results of the median predictor to be selected as the final best point in different sizes of CUs, Pan *et al.* proposed an early termination for the motion estimation process of the quadtree based CTU [8]. However, these algorithms only consider the temporal or spatial correlations, the correlations among different depth CUs and the correlation between the PU mode and the best CTU depth selection are not considered.

In this paper, a fast CTU depth decision algorithm is proposed by considering not only the best depth selection correlation among temporal and spatial neighboring CTUs but also the correlation between the PU mode and the best CTU depth selection. The rest of this paper is organized as follows. The statistical analyses and motivations are presented in Section 2. Then, the details of the proposed fast CTU depth decision algorithm are presented in Section 3. Experimental results are

---

⋆ Corresponding author

This work was supported in part by the Natural Science Foundation of China under Grant 61272289; the Hebei Province Natural Science Foundation under Grant G2010001331; the Colleges and Universities of Hebei Province Science and Technology Research Foundation under Grant QN20131060.

shown in Section 4. At last, Section 5 concludes this paper.

2. MOTIVATIONS AND STATISTICAL ANALYSES

The quadtree structure based CTU is a significant innovation in HEVC. The CTU is the basic unit of coding, which is defined by the Largest Coding Unit (LCU) and hierarchical depth. Fig. 1 shows an example of the quadtree structure based CTU. In HEVC encoding process, a slice is divided into a sequence of CTUs, then the CTU is further partitioned into multiple CUs. Take Fig. 1 as an example, the LCU and hierarchical depth are equal to 64 and 4, respectively. The CTU with quadtree depth 0 can be split into 4 CUs with size of 32×32. Then, the CUs with size of 32×32, which are with quadtree depth 1, are further partitioned into 16 CUs with size of 16×16. Ultimately, the CUs with quadtree depth 2 are divided into 64 CUs with size of 8×8, which are corresponding to quadtree depth 3. The best CTU depth among these four depths is often selected according to the minimization of the Lagrangian cost function [9, 10]

$$m^* = \arg\min_{m \in \mathbf{Q}} D(m) + \lambda \cdot R(m), \tag{1}$$

where $D(m)$ represents the Sum of Squared Differences (SSD) between the original CTU c and its reconstruction c', which is obtained by coding c with depth m; $R(m)$ represents the number of bits which are used for encoding the CTU c with depth m; $\mathbf{Q}$ denotes all quadtree depths, from depth 0 to 3. λ indicates the Lagrange multiplier. The quadtree structure based CTU improves the coding efficiency of HEVC significantly. However, the achieved coding efficiency comes at the cost of high computational complexity of trying all CTU depths. Hence, if the CTU depth decision process can be early terminated, significant encoding time will be saved.

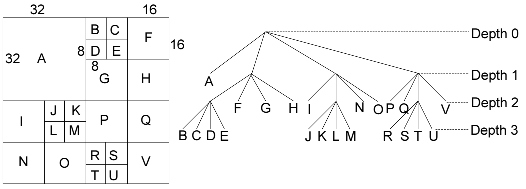

Fig. 1. HEVC CTU partitions based on a quadtree (LCU=64, hierarchical depth=4)

In order to analyze the distribution of the best CTU depth, five video sequences with various resolutions, including *Cactus*, *FourPeople*, *PartyScene*, *SlideShow* and *Traffic* are tested. The experiments are tested under the HEVC common test condition [12], in which Low Delay (LD) configuration is used for *FourPeople*; the other sequences are tested under Random Access (RA) configuration. The LCU and hierarchical depth are equal to 64 and 4, respectively. The Motion Estimation (ME) method and ME search range are fast and 64, respectively. 56 frames to be encoded. The Quantization Parameter (QP) is set to 27. The statistical results are tabulated in Table 1.

Table 1. Probability of the best quadtree depth selection (%)

Resolution	Sequence	Depth 0	Depth 1	Depth 2	Depth3
2560×1600	Traffic	55.43	24.22	13.87	6.48
1920×1080	Cactus	51.62	27.74	14.58	6.06
1280×720	FourPeople	67.06	15.29	15.16	2.50
1280×720	SlideShow	79.34	8.07	10.64	1.95
832×480	PartyScene	24.13	26.56	27.52	21.78
Average		**55.52**	**20.38**	**16.35**	**7.75**

From Table 1, it can be observed that the probability of the best quadtree depth among depth 0, 1, 2 and 3 are different. The quadtree depth 0 holds the probability from 24.13% to 79.34%, 55.52% on average. The quadtree depth 1 and 2 take up the probability on 20.38% and 16.35% on average, respectively. The quadtree depth 3 holds the probability form 1.95% to 21.78%, 7.75% on average. Thus, we can obtain that 1) the quadtree depth 0 holds quite large probability of being selected as the best CTU depth. Hence, if the quadtree depth 0 can be early determined, significant encoding time will be saved; 2) the quadtree depth 3 takes up very small probability, which means quadtree depth 3 is not necessary for most CTUs.

3. PROPOSED FAST CTU QUADTREE DEPTH DECISION ALGORITHM

3.1. Early Quadtree Depth 0 Decision

In video coding process, the content between the current frame and its previously encoded frame may be quite similar. Hence, the best quadtree depth of one CTU in the current frame may be quite similar to the depth of the temporal collocated CTU in its previously encoded frame. In addition, the current CTU has a high coding correlation with its spatial neighboring CTUs. Fig. 2 gives an illustration on the spatial neighboring and temporal collocated CTUs of the current CTU. An experiment is performed to analyze the conditional probability P(**S**|**T**), where **S** represents the event that the current CTU is encoded as quadtree depth 0; **T** denotes the event that the reference CTU is encoded as quadtree depth 0, and **T**={A, B, C, D, E}, they are corresponding to the left, above-left, above, above-right and collocated CTU, respectively. The statistical results are tabulated in Table 2.

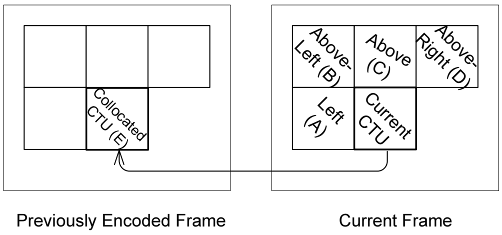

Fig. 2. Neighboring CTUs and its collocated CTU in previously encoded frame

From Table 2, we can see that when the reference CTUs are encoded as quadtree depth 0, the current CTU has a large probability to be encoded as depth 0. When the left and collocated CTUs are encoded as quadtree depth 0, the current CTU has probabilities 0.85 and 0.89 on average to be encoded as quadtree depth 0, respectively. When the above, above-left and above-right CTUs are encoded as quadtree depth 0, the current CTU has probabilities 0.76, 0.74 and 0.73 to be encoded as quadtree depth 0, respectively. Hence, based on the best depth of the reference CTUs, the current CTU which is encoded as quadtree depth 0 can be determined early if

$$\mathrm{W} = \sum_{\mathrm{n} \in \mathbf{T}} \lambda_{\mathrm{n}} \mathrm{C}_{\mathrm{n}} \geq \alpha, \tag{2}$$

where $\mathrm{C_n}$ represent the weight of the reference CTUs, which is 1 if the reference CTU is encoded as depth 0, otherwise, $\mathrm{C_n}$ is equal to 0; λ_n is a weight factor, which is set according to their conditional probability, λ_A=λ_E=1.0, λ_B=λ_C=λ_D=0.75; α is a threshold.

Table 2. Statistical results of the conditional probability P(**S**|**T**)

Sequence	P(**S**|**A**)	P(**S**|**B**)	P(**S**|**C**)	P(**S**|**D**)	P(**S**|**E**)
Traffic	0.82	0.78	0.75	0.77	0.96
Cactus	0.83	0.79	0.75	0.73	0.95
FourPeople	0.88	0.78	0.74	0.76	0.88
SlideShow	0.96	0.87	0.86	0.86	0.94
PartyScene	0.75	0.60	0.60	0.53	0.70
Average	**0.85**	**0.76**	**0.74**	**0.73**	**0.89**

Table 3. Average encoding results for different α

α	BDPSNR (dB)	BDBR (%)	TS (%)
2.75	-0.233	6.16	-48.18
3.25	-0.231	6.04	-48.16
3.50	**-0.013**	**0.27**	**-38.01**
4.25	-0.011	0.25	-32.51

In order to trade off the RD performance and computational complexity saving, a set of α values (2.75, 3.25, 3.50 and 4.25) are tested, where 2.75 represents two direct reference CTUs (left, collocated) and one of three indirect reference CTUs (above-left, above, and above-right) are encoded as quadtree depth 0; 3.25 means one of two direct and three indirect reference CTUs are encoded as quadtree depth 0; 3.50 denotes two direct and two of three indirect reference CTUs are encoded as quadtree depth 0; 4.25 represents all direct and indirect reference CTUs are encoded as quadtree depth 0. Three video sequences *FourPeople*, *PartyScene* and *SlideShow* are used for evaluating the encoding performance in terms of BDPSNR [13], BDBR [13] and total encoding Time Saving (TS). TS is defined as TS $= [(\mathrm{T_p} - \mathrm{T_o})/\mathrm{T_o}] \times 100\%$, where the $\mathrm{T_p}$ and $\mathrm{T_o}$ denote the total encoding time of the proposed method and the original HM8.0 [11], respectively. The test conditions are same as the settings in Section 2. The average test results are tabulated in Table 3. From exhaustive experimental results, the proposed early quardtree depth 0 decision condition can achieve a good trade-off between the encoding complexity reduction and the RD degradation by setting α=3.50.

In order to evaluate the efficiency of the proposed algorithm, Determination Rate (DR) and Hit Rate (HR) are adopted, which are corresponding to the complexity reduction and CTU quadtree depth decision accuracy, respectively, and they are defined as

$$\begin{cases} \mathrm{S_{DR}}(\mathbf{B}|\mathbf{A}) = \mathrm{N}(\mathbf{B}|\mathbf{A})/\mathrm{N}(\mathbf{A}) \times 100\%, \\ \mathrm{T_{HR}}(\mathbf{A}|\mathbf{B}) = \mathrm{N}(\mathbf{A}|\mathbf{B})/\mathrm{N}(\mathbf{B}) \times 100\%, \end{cases} \tag{3}$$

where $\mathrm{S_{DR}}(\mathbf{B}|\mathbf{A})$ and $\mathrm{T_{HR}}(\mathbf{A}|\mathbf{B})$ denote the DR and HR, respectively; $\mathrm{N}(\cdot)$ represents the number of total CTUs of the corresponding event, and the event **A** represents the selected quadtree depth of the encoded CTU, **B** denotes the CTU quadtree depth decision condition. **B**|**A** and **A**|**B** are two conditional events. If DR is large, more coding complexity could be reduced. If HR is large and close to 100%, it means the best CTU quadtree depth is correctly predicted and almost no RD degradation would be caused.

Table 4. DR and HR of the proposed algorithms (%)

Sequence	Early Depth 0 Decision		Depth 3 Skipped Decision	
	DR	HR	DR	HR
Traffic	49.47	98.34	68.38	91.30
Cactus	53.67	97.90	67.28	91.83
FourPeople	70.70	95.56	80.46	95.31
SlideShow	63.58	98.47	63.26	98.16
PartyScene	31.29	91.43	68.11	88.12
Average	**53.74**	**96.34**	**69.50**	**92.94**

To evaluate the efficiency of the proposed early quadtree depth 0 decision condition with α equals to 3.50, the DR and HR defined in Eq. (3) are used, where **A** is the event that the current CTU selects the depth 0 as its best depth, i.e. $\mathrm{D_{best}} = 0$; **B** represents the early depth 0 decision condition, which is defined in Eq. (2) and denoted as $\mathrm{W} \geq \alpha$. The detailed results of $\mathrm{S_{DR}}(\mathrm{W} \geq \alpha|\mathrm{D_{best}} = 0)$ and $\mathrm{T_{HR}}(\mathrm{D_{best}} = 0|\mathrm{W} \geq \alpha)$ are listed in Table 4. From Table 4, it can be observed that there are about 53.74% CTUs which select quadtree depth 0 as their best depth can be early determined by using the proposed method. The HR holds 96.34% on average. In other words, 96.34% CTUs which select quadtree depth 0 as their depth can be correctly determined. These values demonstrate that the proposed quadtree depth 0 early determination algorithm works efficiently.

3.2. Quadtree Depth 3 Skipped Decision

PU is the unit of inter/intra prediction and a single CU can contain multiple PUs. One innovation of PU is that asymmetric partitions of a CU for inter prediction are used. When the asymmetric PU mode is selected as the best mode in PU encoding process, it represents the current CU is an irregular image pattern, and this CTU has a large probability to be encoded in smaller size CUs. On the other hand, if the INTRA mode is selected as the best mode in PU coding process, it represents this CU moves fast or with complex content, and inter prediction can not locate a best matching CU in the reference frame. In this case, this CTU also has a large probability to be encoded as smaller size CUs. In addition, quadtree depth 3 has the smallest percentage to be chosen as the best depth according to Table 1. Hence, quadtree depth 3 is most likely performed only when

$$\mathrm{P}_1 \in \mathbf{M} \;\&\&\; \mathrm{P}_2 \in \mathbf{M}, \tag{4}$$

where $\mathbf{M}$={2N×nU, 2N×nD, nL×2N, nR×2N, INTRA}, P_1 and P_2 denote the best PU mode in CTU with quadtree depth 1 and 2, respectively. && is an and operation, which means the quadtree depth 3 is performed for the current CTU when it can simultaneously meet these two requirements, $\mathrm{P}_1 \in \mathbf{M}$ and $\mathrm{P}_2 \in \mathbf{M}$.

While testing the efficiency of the proposed quadtree depth 3 skipped decision method, the DR and HR defined in Eq. (3) are adopted, **A** is the event that the current CTU selects the best quadtree depth from 0, 1 and 2, which is denoted as $\mathrm{D}_{\mathrm{best}} \neq 3$. **B** represents the event that the depth 3 skipped decision condition, which is $\mathrm{P}_1 \in \mathbf{M}\&\&\mathrm{P}_2 \in \mathbf{M}$. The detailed results of $\mathrm{S}_{\mathrm{DR}}(\mathrm{P}_1 \in \mathbf{M}\&\&\mathrm{P}_2 \in \mathbf{M}|\mathrm{D}_{\mathrm{best}} \neq 3)$ and $\mathrm{T}_{\mathrm{HR}}(\mathrm{D}_{\mathrm{best}} \neq 3|\mathrm{P}_1 \in \mathbf{M}\&\&\mathrm{P}_2 \in \mathbf{M})$ are tabulated in Table 4. From Table 4, it can be observed that there are about 69.50% CTUs which don't select quadtree depth 3 as their best depth can be early determined by using the proposed depth 3 skipped decision condition. The HR holds 92.94% on average. In other words, the accuracy is 92.94% and very small number of CTUs will be wrongly determined. These values demonstrate that the proposed quadtree depth 3 skipped decision algorithm works efficiently.

3.3. The Overall Algorithm

Based on above analyses, the proposed fast CTU depth decision algorithm is summarized and illustrated step-by-step as follows.

Step 1. Encode the current CTU with quadtree depth 0, if Eq. (2) is satisfied, go to Step 4; otherwise, go to Step 2.

Step 2. Encode the current CTU with quadtree depths 1 and 2, if Eq. (4) is satisfied, go to Step 3; otherwise, go to Step 4.

Step 3. Encode the current CTU with quadtree depth 3. Go to Step 4.

Step 4. Store the coding information and choose the best quadtree depth among all tested depths based on Eq. (1). Go back to Step 1 to encode the next CTU.

4. EXPERIMENTAL RESULTS

In order to evaluate the efficiency of the proposed algorithm, HEVC reference software HM8.0 is used for the software platform. Sixteen video sequences are encoded under HEVC common test condition, the LD configuration is used for 1280×720 video sequences. Four QPs (22, 27, 32 and 37) are tested. The other test conditions are same as the settings in Section 2. The hardware platform is Intel Core 2 Duo CPU E5800 @ 3.16GHz and 3.17GHz, 4.00GB RAM with Microsoft Windows 7 64-bit operating system.

Table 5. Summary of encoding results

Resolution	Sequence	BDPSNR (dB)	BDBR (%)	TS (%)
2560×1600	Traffic	-0.012	0.35	-42.96
	PeopleOnStreet	-0.018	0.48	-28.88
1920×1080	Kimono	-0.025	0.78	-35.19
	ParkScene	-0.024	0.79	-37.58
	Cautus	-0.012	0.60	-38.38
	BQTerrace	-0.019	0.98	-41.44
1280×720	Johnny	-0.016	0.56	-53.26
	KristernAndSara	-0.011	0.35	-52.00
	Vidyo1	-0.008	0.26	-49.54
	Vidyo3	-0.035	0.97	-47.37
	Vidyo4	-0.017	0.73	-48.63
832×480	BQMall	-0.035	0.84	-30.76
	PartyScene	-0.024	0.51	-27.94
	Flowervase	-0.020	0.57	-45.46
416×240	BQSquare	-0.016	0.38	-24.03
	Mobisode2	-0.030	0.62	-34.78
Average		**-0.020**	**0.61**	**-40.01**

The coding performance of the proposed algorithm is compared with the original HM8.0 in terms of BDPSNR, BDBR and TS. The experimental results are compared and summarized in Table 5. From Table 5, it can be observed that the proposed algorithm can reduce the computational complexity from 24.03% to 53.26%, 40.01% on average. The BDPSNR between the proposed algorithm and the original HM8.0 is from -0.008 to -0.035 dB, -0.018 dB on average. The BDBR between the proposed algorithm and the original HM8.0 is from 0.26% to 0.98%, 0.61% on average. For all 1280×720 sequences and *Traffic*, the proposed algorithm can significantly reduce the computational complexity, this is because these sequences have a large number of static regions which are quite suitable for encoding as quadtree depth 0. For video sequences with smaller resolution (*BQSquare*) or with fast motion activity and complex content (*PartyScene*, *PeopleOnStreet*), only about 25% encoding time can be saved by the proposed algorithm. The reason is that these sequences are with fast and complex content, and are suitable for encoding with smaller size CUs.

5. CONCLUSION

In this paper, we propose a fast CTU depth decision algorithm for HEVC. Firstly, according to the best quadtree depth of the spatial and temporal neighboring CTUs of the current CTU, an early quadtree depth 0 decision algorithm is proposed. Then, based on the correlations between the PU mode and the best CTU depth selection, a quadtree depth 3 skipped decision algorithm is proposed. Experimental results demonstrate that the proposed algorithm can work efficiently in reducing the computational complexity and maintaining a comparable RD performance.

6. REFERENCES

[1] G.J. Sullivan, J. Ohm, W.-J. Han, T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 22, no. 12, pp.1649-1668, Dec. 2012.

[2] ITU-T and ISO/IEC JCT-VC, "Infrastructure of audio-visual services-coding of moving video," ITU-T Rec. H.265, Apr. 2013.

[3] K. Choi, S.-H. Park, E.S. Jang, "Coding tree pruning based CU early termination," ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) document JCTVC-F092, Jul. 2011.

[4] X. Li, J. An, X. Guo, S. Lei, "Adaptive CU depth range," ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) document JCTVC-E090, Mar. 2011.

[5] J. Kim, S. Jeong, S. Cho, J.-S. Choi, "Adaptive coding unit early termination algorithm for HEVC," in Proceedings International Conference on Consumer Electronics (ICCE), IEEE, pp. 261-262, Jan. 2012.

[6] X. Shen, L. Yu, J. Chen, "Fast coding unit size selection for HEVC based on Bayesian decision rule,"in Proceedings Picture Coding Symposium (PCS), IEEE, pp.453-456, May 2012.

[7] L. Shen, Z. Liu, X. Zhang, W. Zhao, Z. Zhang, "An effective CU size decision method for HEVC encoders," *IEEE Trans. on Multimedia*, vol.15, no.2, pp.465-470, Feb. 2013.

[8] Z. Pan, Y. Zhang, S. Kwong, X. Wang, L. Xu, "Early termination for TZSearch in HEVC motion estimation," in proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1389-1393, May 2013.

[9] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan,"Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 13, no. 7, pp. 688-703, Jul. 2003.

[10] J. Ohm, G.J. Sullivan, H. Schwarz, T.K. Tan; T. Wiegand, "Comparison of the coding efficiency of video coding standards-including High Efficiency Video Coding (HEVC)," *IEEE Trans. Circuits Syst. Video Techn.*, vol.22, no.12, pp.1669-1684, Dec. 2012.

[11] B. Bross, W.-J. Han, J.-R. Ohm, G. Sullivan, T. Wiegand, "High efficiency video coding (HEVC) text specification draft 8," ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) document JCTVC-J1003, Jul. 2012.

[12] F. Bossen, "Common test conditions and software reference configurations," ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) document JCTVC-J1100, Jul. 2012.

[13] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," ITU-T VCEG, document VCEG-M33, Austin, TX, Apr. 2001.

978-1-4799-5751-4/14/$31.00 ©2014 IEEE