Prediction Mode Based Reference Line Synthesis for Intra Prediction of Video Coding

Qiang Yao
KDDI Research, Inc. Ohara 2-1-15
Fujimino, Saitama, Japan
Email: qi-yao@kddi-research.jp

Kei Kawamura
KDDI Research, Inc. Ohara 2-1-15
Fujimino, Saitama, Japan
Email: kei@kddi-research.jp

Sei Naito
KDDI Research, Inc. Ohara 2-1-15
Fujimino, Saitama, Japan.
Email: sei@kddi-research.jp

***Abstract*—Intra prediction is a significant coding tool that allows a high level of video compression to be achieved in the current state-of-the-art video coding standard, High Efficiency Video Coding (HEVC), and the joint exploration model (JEM) developed by the Joint Video Exploration Team (JVET) of ITU-T VCEG and ISO/IEC MPEG for the next generation video coding standard. In intra prediction, the top and left adjacent lines to the current coding block in the neighboring reconstructed blocks are selected as the reference lines. However, it has been observed that the adjacent reference line might not always provide optimal prediction due to the quantization noise and object occlusions caused by straight lines.**

In this paper, we propose the synthesis of another reference line by integrating multiple lines in the neighboring reconstructed blocks based on the prediction mode. The synthesized line and the reconstructed adjacent line compete in the rate distortion optimization process, where the line that yields a minimum cost is finally selected. The proposed method is implemented on top of JEM 3.0, and the experimental results show that -0.29% (average), -1.15% (maximum) luma BD gain and -0.15% (average), -0.59% (maximum) luma BD gain can be achieved in all intra and random access conditions, respectively, among all the test sequences.

1. Introduction

Intra prediction is a key technique for achieving high compression performance by reducing spatial redundancy among samples, and it is a significant coding tool in both the HEVC test model (HM) for the current state-of-the-art video coding standard, High Efficiency Video Coding (HEVC) [1], and the joint exploration model (JEM) [2] developed on top of HM16.6 by the Joint Video Exploration Team (JVET) [3] of ITU-T VCEG and ISO/IEC MPEG for the next generation video coding standard. In HEVC intra prediction [4], there are three main steps: reference sample generation, target sample prediction, and the filtering process. The residual between the target sample and the corresponding predicted sample is then transformed, quantized and further coded into bitstream. Intra picture accounts for a large portion of bitstream, and, as the reference of the following inter picture, the quality of the intra picture also affects the coding performance of the following inter pictures. Thus, improvement of intra prediction contributes to coding performance for both still pictures and motion pictures in video sequences.

The intra prediction structure in JEM is inherited and modified from the one in HM. Besides keeping DC mode and planar mode, the number of angular intra mode in JEM has been extended from 33 to 65, and denser prediction angle brings about higher prediction accuracy. Other modifications in the intra prediction of JEM can be found in the algorithm description document [6]. However, there is no change between HM and JEM in the selection of the reference line for intra prediction. The assumption of intra prediction is that the prediction accuracy is inversely proportional to the spatial distance between the current sample and the reference sample. However, even if the nearest reference line has the strongest spatial correlation to the samples in the current block, such reference line lies in the farthest position to its own reference line in the neighboring block. In other words, when the quantization step is large, the quantization error of the residual becomes large, and hence the adjacent reference line has probably suffered the most degradation due to the large quantization step in the previous prediction process. Moreover, when there is a thin object, e.g. a stick or a straight line, occluding the background, a more appropriate partitioning structure can be obtained if another synthesized line is available as the reference line.

Therefore, in this paper, we propose to synthesize another reference line by integrating multiple lines in the reconstructed neighboring blocks based on the prediction mode. There is a competition between the synthesized reference line and the adjacent reference line in the rate distortion optimization (RDO) process, and the line that yields the minimum RDO cost is finally selected. Experimental results show that on average a -0.29% and -0.15% luma BD gain can be achieved in all intra and random access conditions, respectively, and the maximum gain reaches -1.15% and -0.59% in all intra and random access conditions, respectively.

The remainder of the paper is organized as follows. Related work is introduced in the next section, and the proposed method is described in section 3, followed by a

ISBN 978-0-9928626-7-1 © EURASIP 2017

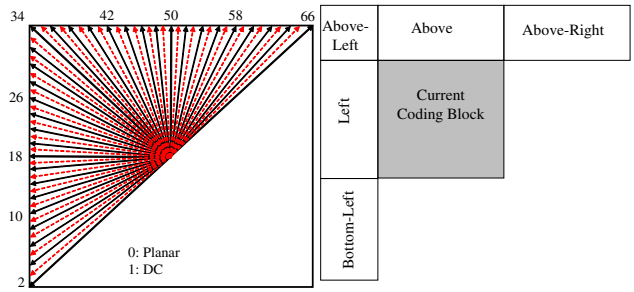
description of the experimental results in section 4. Finally, we briefly conclude this paper in the last section.

2. Related work

Since the release of HEVC, there have been a number of studies on the improvement of intra prediction. Y. Chen et al. viewed the image signal as a 2D non-separable Markov model and proposed the use of three-tap extrapolation filters as a replacement for the pixel-copying prediction mode [7]. However, this kind of recursive extrapolation approach reduced the parallelism of the coding process. Later, F. Kamisli proposed the adoption of a 2D Markov process for both intra prediction and transform step in order to achieve improved coding gain [8]. Furthermore, a sparse coding scheme was also proposed by L.F. Lucas et al. to generate prediction using sparse linear predictors, and geometric transformation was also considered in the generation of prediction [9]. The schemes mentioned above challenged the traditional structure of intra prediction, and explored the statistical, geometric, and sparse characteristics of image content in a quest for more accurate prediction. However, a major concern is the difficulty to fully support parallel computing, which has become very popular in hardware implementation.

On the other hand, there are also several studies on intra prediction that follow the structure in HEVC. X. Chao et al. proposed a short distance intra coding method by splitting a coding unit (CU) into non-square prediction units (PUs) so that the shortened spatial distance between the current sample and the reference sample brings higher prediction accuracy [10]. The idea has been absorbed into a new partition structure, called quad tree plus binary tree (QTBT) [6] in JEM. In addition, X. Qi et al. proposed an intra prediction method based on inpainting algorithms and vector prediction [11]. Based on the assumption that repetitive or similar patches may exist over a natural image, a vector is generated by using template matching in a reconstructed search region on top and left of the current block. However, the reconstruction of a search region leads to greater memory access in the implementation. Moreover, J. Li et al. proposed trying multiple reference lines in RDO process and selecting the best one [12]. However, each line is simply tried as the reference line without any further processing. Furthermore, E. Wige et al. proposed a pixel-wise prediction based on original samples for lossless coding [13]. The causal samples of the current sample are reconstructed and grouped into several patterns as the reference in prediction. However, the dependency of samples reduces the parallelism of the entire prediction.

X. Chao, X. Qi, J. Li, and E. Wige et al. tried to explore different reference samples or reference lines in the intra prediction [10]–[13], and our proposed method shares a similar intention but differs substantially in terms of its essence. We propose to further synthesize a reference line to provide an alternative when the nearest reference line fails to provide a good prediction. In addition, the angular information is utilized to robustly synthesize the reference



(a) Angular prediction (b) Neighboring blocks

Figure 1. Illustration of angular intra prediction in JEM

line. As far as we know, it is the first time a reference line for intra prediction has been synthesized, and this addition of a synthesized reference line brings -0.29% and -0.15% averaged BD gain in all intra and random access conditions respectively.

3. Proposed Method

3.1. Overview

An illustration of the intra prediction mode in JEM is shown in **Figure 1 (a)**, where red dotted arrows represent added finer prediction angles. A two-step intra mode candidate construction for RDO is performed, where the coarse 35 modes are checked based on the sum of absolute transformed differences (SATD) of the prediction residual to select the best N modes in the first step and then the neighboring finer angles of the selected modes are further checked by SATD to update the best N modes in the second step. Next, the first M most probable modes (MPMs) that have been derived from the top and left neighboring PUs are combined with the N selected modes to form the candidate list of intra prediction mode in the RDO process. In addition, as shown in **Figure 1 (b)**, the reference line selection in both HM and JEM is the same.

However, when there is a thin object, e.g. a stick or a straight line, occluding the background or there is large quantization error in the adjacent reference line, a more appropriate partitioning structure can be obtained if another synthesized line is available as the reference line.

Therefore, in order to enhance the robustness and accuracy of intra prediction, we propose to synthesize another reference line by using multiple lines in the neighboring block based on the prediction mode, and the basic flow of the proposed method is shown in **Figure 2**, where the bold boxes indicate the main parts of our proposal. The synthesized reference line is checked in the RDO process after the construction of a candidate list of intra prediction mode, and a flag in the prediction unit (PU) level is added into bitstream to indicate the decoder which reference line is finally adopted in the prediction. Please note that the proposed method is only for the luma component.

ISBN 978-0-9928626-7-1 © EURASIP 2017

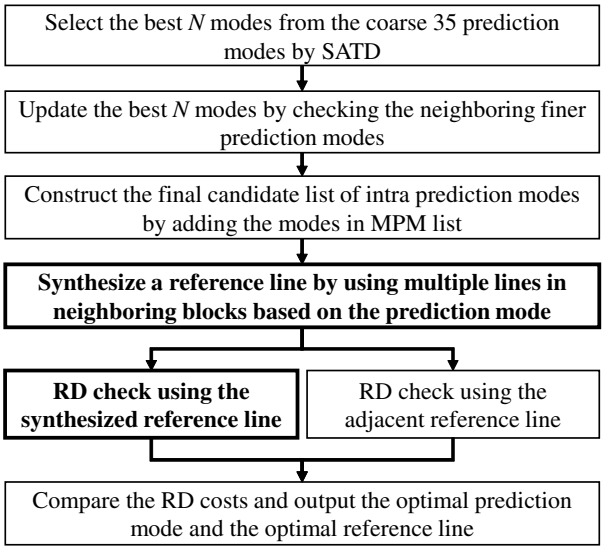


Figure 2. The basic procedure of our proposal (The bold boxes are the main part in our proposal)

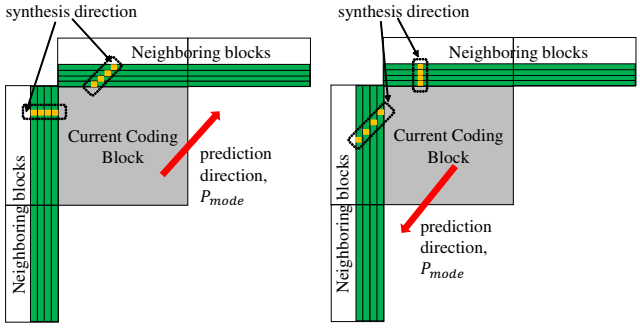


Figure 3. The illustration of reference line synthesis

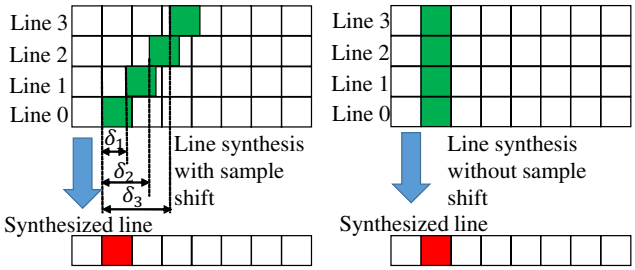


Figure 4. The illustration of sample shift in synthesis of a top reference line

3.2. Prediction Mode-based Reference Line Synthesis

Multiple lines in the neighboring block are adopted in the reference line synthesis, and the number of utilized multiple lines is limited to four in order to reduce the memory access. Another reason for selecting four multiple lines is that the correlation between reference lines and the current coding block drops drastically if more distant lines are utilized.

Let L^t_{syn} and L^l_{syn} denote the top and the left synthesized reference line respectively to the current PU. In addition, we define three categories of prediction mode: *horizontal mode* ($1 < P_{mode} \leq 34$), *vertical mode* ($34 < P_{mode} \leq 66$), and *non-angular mode* ($P_{mode} = 0$; Planar, or $P_{mode} = 1$; DC). Thus, given a prediction mode, P_{mode}, we first check the category of the mode. As shown in **Figure 3**, if the prediction mode P_{mode} belongs to the *horizontal mode*, the synthesis of L^l_{syn} follows the P_{mode} direction and the synthesis of L^t_{syn} follows the direction of $P_{mode} = 50$. Symmetrically, if P_{mode} belongs to *vertical mode*, only the synthesis of L^t_{syn} follows the direction of P_{mode} and L^l_{syn} is synthesized by following the direction of $P_{mode} = 18$. Moreover, if P_{mode} is the *non-angular mode*, the left and top reference lines are synthesized by following the direction of $P_{mode} = 18$ and $P_{mode} = 50$, respectively.

In the next step, the specific synthesis method is explained. The sample in the synthesized line and reconstructed line is defined as $p_{syn}[\bullet][\bullet]$ and $p_{rec}[\bullet][\bullet]$ respectively. The coordinate of the top-left sample in the current block is defined as $[0][0]$. Thus, the adjacent sample to the current block in the top-left neighboring block is defined as $p_{rec}[-1][-1]$, and there is

$$p_{syn}[-1][-1] = \frac{1}{8}\sum_{i=-4}^{-1}(p_{rec}[i][-1] + p_{rec}[-1][i]). \quad (1)$$

As for the line synthesis, the sample in the synthesized line is calculated by averaging the four corresponding samples in the four reference lines, as shown in **Figure 3**. Therefore, the synthesis of the top reference line, L^t_{syn}, is written as

$$p_{syn}[x][-1] = \frac{1}{4}\sum_{i=0}^{3} p_{rec}[x][-1-i], \quad (2)$$

where $x \in [0, W-1]$ and W is the length of the top reference line. Similarly, the left reference line, L^l_{syn}, can be synthesized as

$$p_{syn}[-1][y] = \frac{1}{4}\sum_{j=0}^{3} p_{rec}[-1-j][y], \quad (3)$$

where $y \in [0, H-1]$ and H is the length of the left reference line.

Since the prediction direction has been available before the reference line is synthesized, it is important and necessary to synthesize the reference line along such direction, otherwise uncorrelated samples are involved in the reference line synthesis, resulting in more prediction errors. In order to calculate an appropriate shift for each reference line, we propose to reuse the modified sample prediction scheme. As shown in **Figure 4**, a displacement value $\alpha \in [-32, 32]$ for

ISBN 978-0-9928626-7-1 © EURASIP 2017

TABLE 1. THE SUMMARY OF BD RATE AND RUMTIME (ALL INTRA)

All Intra Main10						
Class	Sequence Name	Y	U	V	Enc T	Dec T
A1	Tango	-0.32%	-0.06%	0.12%	183%	101%
	Drums100	-0.24%	0.00%	-0.02%	186%	101%
	CampfireParty	-0.25%	0.04%	0.04%	185%	101%
	ToddlerFountain	-0.01%	-0.01%	0.02%	189%	101%
A2	CatRobot	-0.33%	0.08%	0.00%	185%	100%
	TrafficFlow	**-1.15%**	-0.56%	-0.54%	183%	100%
	DaylightRoad	-0.65%	-0.13%	-0.10%	186%	101%
	Rollercoaster	-0.24%	0.02%	0.06%	179%	101%
B	Kimono	-0.07%	0.21%	-0.01%	185%	101%
	ParkScene	-0.05%	-0.08%	0.02%	189%	100%
	Cactus	-0.18%	-0.05%	0.00%	188%	101%
	BasketballDrive	-0.39%	-0.02%	-0.08%	185%	101%
	BQTerrace	-0.79%	-0.43%	-0.54%	188%	100%
C	BasketballDrill	-0.61%	-0.02%	-0.36%	188%	100%
	BQMall	-0.15%	-0.16%	-0.03%	188%	100%
	PartyScene	-0.17%	0.03%	0.14%	191%	100%
	RaceHorsesC	-0.13%	-0.02%	0.06%	189%	101%
D	BasketballPass	-0.09%	-0.20%	0.08%	187%	102%
	BQSquare	-0.21%	0.08%	-0.09%	192%	101%
	BlowingBubbles	-0.14%	-0.08%	0.22%	191%	100%
	RaceHorses	-0.09%	0.12%	0.17%	191%	101%
E	FourPeople	-0.11%	0.09%	-0.07%	187%	101%
	Johnny	-0.46%	-0.23%	-0.18%	183%	100%
	KristenAndSara	-0.21%	0.16%	-0.05%	185%	100%
	Overall	**-0.29%**	-0.05%	-0.05%	187%	101%

each reference line is obtained at 1/32 pixel accuracy from a look up table according to a given prediction mode P_{mode}, and the α is further converted to a different sample offset δ_d by using

$$\delta_d = \begin{cases} \left\lfloor \dfrac{\alpha * d}{32} \right\rfloor, & \text{if } (\alpha * d) \text{ mode } 32 \leq 15 \\ \left\lfloor \dfrac{\alpha * d}{32} \right\rfloor + 1, & \text{if } (\alpha * d) \text{ mode } 32 > 15, \end{cases} \tag{4}$$

where $d \in [0, 3]$ is the distance from the reference lines to the current block. Consequently, when the prediction direction is considered in the reference line synthesis, the **eq. (2)** and **eq. (3)** are modified as

$$p_{syn}[x][-1] = \frac{1}{4}\sum_{i=0}^{3} p_{rec}[x + \delta_i][-1 - i], \tag{5}$$

$$p_{syn}[-1][y] = \frac{1}{4}\sum_{j=0}^{3} p_{rec}[-1 - j][y + \delta_j], \tag{6}$$

Please note **eq. (4)** outputs an integer offset according to the displacement value α and the distance d. Actually, a fractional linear interpolation between two neighboring samples was also tested but was not considered, because the interpolation increases computation complexity in both the encoder and decoder with negligible improvement of coding gain.

4. Experimental Result

The proposed method is implemented on top of JEM 3.0. By following the common test condition [14] in JVET, the proposed method is tested for all 24 sequences (class A1: 4096x2160; class A2: 3840x2160; class B: 1920x1080; class C: 832x480; class D: 416x240; class E: 1280x720) in all intra and random access conditions, respectively.

TABLE 2. THE SUMMARY OF BD RATE AND RUNTIME (RANDOM ACCESS)

Random Access						
Class	Sequence Name	Y	U	V	Enc T	Dec T
A1	Tango	-0.26%	-0.46%	-0.31%	120%	100%
	Drums100	-0.09%	0.10%	0.16%	115%	99%
	CampfireParty	-0.22%	0.09%	-0.07%	133%	100%
	ToddlerFountain	-0.02%	-0.20%	-0.04%	139%	99%
A2	CatRobot	-0.16%	-0.06%	0.07%	115%	98%
	TrafficFlow	**-0.59%**	-0.09%	-0.01%	121%	99%
	DaylightRoad	-0.27%	0.27%	0.38%	119%	100%
	Rollercoaster	-0.01%	0.06%	0.21%	116%	99%
B	Kimono	-0.03%	-0.23%	-0.14%	117%	99%
	ParkScene	0.01%	0.26%	-0.05%	121%	100%
	Cactus	-0.05%	-0.01%	-0.10%	120%	100%
	BasketballDrive	-0.07%	-0.33%	0.01%	124%	100%
	BQTerrace	-0.43%	0.13%	-0.07%	112%	99%
C	BasketballDrill	-0.39%	0.19%	0.05%	118%	99%
	BQMall	-0.12%	-0.14%	-0.13%	115%	99%
	PartyScene	-0.10%	0.00%	-0.22%	117%	99%
	RaceHorsesC	-0.16%	-0.07%	0.20%	122%	99%
D	BasketballPass	-0.03%	0.89%	-0.03%	120%	99%
	BQSquare	-0.09%	0.17%	-0.36%	112%	98%
	BlowingBubbles	0.05%	-0.67%	0.05%	114%	99%
	RaceHorses	-0.06%	0.46%	-0.07%	119%	99%
	Overall	**-0.15%**	0.02%	-0.03%	119%	99%

The summary of the BD rate [15] and runtime are shown in **Table 1** and **Table 2**. **Table 1** shows the results in all intra condition, and it is observed that an average -0.29% and up to -1.15% luma BD rate gain is achieved with an 87% average runtime increase on encoder side. In addition, according to the results in random access condition shown in **Table 2**, an average -0.15% and up to -0.59% luma BD rate gain is achieved with a 19% average runtime increase on encoder side. For all intra condition, there is a 1% decoder runtime increase, and there is no decoder runtime increase in random access condition. From the two tables, it can be seen that the sequences 'TrafficFlow' in class A2 and 'BQTerrace' in class B are the sequences most improved by using the proposed method. The main reason is that there are many thin and straight lines as textures in the both sequences, and hence the synthesized reference line could provide a better prediction than the adjacent reference line. Furthermore, it is observed that there is almost no improvement to the sequence 'ToddlerFountain' in class A1 because of the complicated content that the sprayed water is splashed in various directions.

Finally, a subjective comparison is made between the original picture, the coded picture by JEM 3.0, and the coded picture by the proposed method. 'TrafficFlow' sequence is selected and encoded at the QP value of 37 in all intra condition. One picture is selected and the cropped part (in the resolution of 300x200) is compared in **Figure 5**, and it is observed that the visual quality is also improved by the proposed method.

5. Conclusion

This paper proposed the synthesis of a reference line using multiple lines in the neighboring blocks for intra

ISBN 978-0-9928626-7-1 © EURASIP 2017

(a) Original picture (b) Coded by JEM 3.0 (c) Coded by the proposal

Figure 5. Subjective comparison between the original picture, the decoded pictures by JEM 3.0 and our proposed method. (The sequence 'TrafficFlow' is encoded at the QP value of 37 in all intra condition, and the cropped pictures in the resolution of 300x200 are illustrated for detail comparison)

prediction based on the prediction mode. The proposed method was implemented on top of JEM 3.0. In all the test sequences, an average -0.29% and -0.15% luma BD gain in average was achieved in all intra and random access conditions, respectively, and the maximum gain reached -1.15% and -0.59% in all intra and random access condition, respectively. In the near future, we plan to further improve the coding performance and to design fast algorithms for reducing encoder runtime.

References

[1] G. J. Sullivan ,J.-R. 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, 2012.

[2] https://jvet.hhi.fraunhofer.de/svn/svn_HMJEMSoftware/tags/HM-16.6-JEM-3.0/

[3] G. J. Sullivan and J.-R. Ohm, "Meeting Report of the 3rd JVET Meeting", Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-C1000, May, 2016.

[4] J. Lainema, F. Bossen, W.-J. Han, J. Min and K. Ugur, "Intra coding of the HEVC standard", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 22, No. 12, pp. 1792–1801, 2012.

[5] K. Uger and J. Lainema, "Updated results on HEVC still picture coding performance", Joint Collaborative Team on Video Coding, Document JCTVC-M0041, Apr. 2013.

[6] J. Chen, E. Alshina, G. J. Sullivan, J.-R. Ohm, and J. Boyce, "Algorithm Description of Joint Exploration Test Model 3", Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-C1001, May, 2016.

[7] Y. Chen, J. Han, and K. Rose, "A recursive extrapolation approach to intra prediction in video coding", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* pp. 1734–1738, 2013.

[8] F. Kamisli, "Block-based spatial prediction and transforms based on 2D Markov processes for image and video compression", *IEEE Transactions on Image Processing,* Vol. 24, No. 4, pp. 1247–1260, 2015.

[9] L.-F. Lucas, N.-M. Rodrigues, E.-A. da Silva, C.-L. Pagliari, and S.-M. de Faria, "Image Coding Using Generalized Predictors Based on Sparsity and Geometric Transformations", *IEEE Transactions on Image Processing,* Vol. 25, No. 9, pp. 4046–4060, 2016.

[10] X. Cao, C. Lai, Y. Wang, and Y. He, "Short distance intra coding scheme for HEVC", *Picture Coding Symposium (PCS)*, pp. 501–504, 2012.

[11] X. Qi, T. Zhang, F. Ye, A. Men, and B. Yang, "Intra prediction with enhanced inpainting method and vector predictor for HEVC", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1217–1220, 2012.

[12] J. Li, B. Li, J. Xu, and R. Xiong, "Multiple Line-based Intra Prediction for High Efficiency Video Coding", arXiv preprint arXiv:1605.08308, 2016.

[13] E. Wige, G. Yammine, P. Amon, A. Hutter, and A. Kaup, "Pixel-based averaging predictor for HEVC lossless coding", *IEEE International Conference on Image Processing (ICIP),* pp. 1806–1810, 2013.

[14] K. Sühring, and X. Li, "JVET common test conditions and software reference configurations", Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-B1010, Feb. 2016.

[15] G. Bjøntegarrd, "Calculation of Average PSNR Difference between RD-curve", ITU-T Q6/SC16 13th VCEG meeting, VCEG-M33, 2001.

ISBN 978-0-9928626-7-1 © EURASIP 2017