Screen Content Coding for HEVC Using Edge Modes

Hu, S.; Cohen, R.A.; Vetro, A.; Kuo, C.C.J.

TR2013-034 May 2013

Abstract

In this work, a new coding tool called Edge Mode is proposed for HEVC intra coding, aimed at improving coding efficiency for screen content video. A set of edge modes that correspond to edge positions are identified based upon intra prediction directions. Then, a simplified scheme is developed to select the best edge mode. To avoid applying a transform over strong edges, directional 2D separable transforms are applied to blocks partitioned using these edge modes. Experimental results show that HEVC with edge modes (HEVC/EM) can achieve up to an 18% reduction in bit-rate as compared to unmodified HEVC, with an average reduction of 10.4% for screen content video sequences.

IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

SCREEN CONTENT CODING FOR HEVC USING EDGE MODES

Sudeng Hu[1†] *Robert A. Cohen*[2] *Anthony Vetro*[2] *C.-C. Jay Kuo*[1]

[1]Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA
[2]Mitsubishi Electric Research Laboratories, Cambridge, MA

ABSTRACT

In this work, a new coding tool called Edge Mode is proposed for HEVC intra coding, aimed at improving coding efficiency for screen content video. A set of edge modes that correspond to edge positions are identified based upon intra prediction directions. Then, a simplified scheme is developed to select the best edge mode. To avoid applying a transform over strong edges, directional 2D separable transforms are applied to blocks partitioned using these edge modes. Experimental results show that HEVC with edge modes (HEVC/EM) can achieve up to an 18% reduction in bit-rate as compared to unmodified HEVC, with an average reduction of 10.4% for screen content video sequences.

Index Terms— Screen content video coding, HEVC

1. INTRODUCTION

Due to rapidly growing video applications in areas such as wireless display and cloud computing [1], screen content coding has received much interest from academia and industry in recent years [2]-[8]. The High Efficiency Video Coding (HEVC) standard [9] has achieved significant improvement in coding efficiency as compared with the state-of-the-art H.264/AVC standard [10]. However, HEVC has been designed mainly for natural video captured by cameras. Screen content images and video, also known as compound images, hybrid images, and mixed-raster content material, typically contains computer-generated content such as text and graphics, sometimes in combination with natural or camera-captured material. Since the properties of screen content are quite different from those of natural content, and HEVC currently does not exploit these properties, there is still room for improvement in coding efficiency.

There has been a lot of research done on the classification of screen and natural content, e.g., [11]-[14]. Since screen content video may contain artificial content generated by computers, it tends to have sharp edges on object boundaries. The strong edges will lead to discontinuities in the residual signal after intra prediction, and these discontinuities will spread the energy over a wide frequency range, thus reducing the efficiency of transform-based coders such as HEVC. To address this issue, a new intra mode called residual scalar quantization was proposed in [15], where the residual signal is directly encoded by an entropy coder without performing the DCT transform. A similar transform skip was proposed in [16], where the 2D transform can be skipped in either one or both directions. A method was proposed in [17] to quantize residual signals adaptively in either the transform or the spatial domain. These papers report improvements in coding efficiency by skipping the transform for some blocks.

[†]This work was done while at MERL

For screen content, it is our observation that directly encoding residual signals in the spatial domain may not be efficient enough. This is because, except for the edge, the remaining areas are still smooth and can be coded more effectively with a transform. In this paper, we propose a new scheme, called *Edge Mode* (EM), to encode these kinds of blocks. Based on the intra prediction direction, six possible edge positions inside a block are defined, and one of them will be selected via rate-distortion (RD) optimization. To reduce the encoding complexity, the proposed scheme can be further simplified by classifying intra modes into four categories. Then, $M \times N$ 2D DCT transforms or non-orthogonal 2D transforms are performed separately in sub-blocks. Finally, the new edge mode is integrated into HEVC to result in a more powerful coding scheme. This coding scheme is described in Section 2. Experimental results for this method integrated into HEVC are discussed in Section 3, and conclusions are given in Section 4.

2. HEVC EDGE MODE (EM) SCHEME

2.1. Overview of concept

For HEVC intra modes, the content of a square block, or more specifically a prediction block, is predicted using modes having different prediction angles. The mode yielding the least distortion, typically measured as mean-square error or mean absolute error, is selected to code the associated prediction block. Usually, regions with complex content are likely to be coded with a smaller block size such as 4×4 or 8×8, because prediction over a larger block would generate residuals having high energy. For screen content, many of the blocks contain smooth areas separated by a straight or curved edge. There are common methods outside of HEVC to represent such curves using multiscale straight or curved kernels, such as ridgelets and curvelets [18]. Within the HEVC framework, it is therefore reasonable to approximate edges as straight lines with different orientations in blocks of such small sizes. Furthermore, the prediction mode oriented along the edge is likely to produce less residual energy than a mode that predicts across the edge, as pixel values from neighboring blocks used during the prediction process are not good predictors of pixels on the opposite side of an edge. Consequently, if we have selected the best intra prediction for a given block, it is likely that the edge orientation is in parallel with the intra prediction direction. The remaining unknown is the position of the edge within the block.

Along the edge direction, we consider three different edge positions as illustrated in Fig. 1: one passing through the center of block and the other two having a distance of $\frac{1}{4}d\csc\theta$ with respect to the center, where d is the block width and θ is the angle between the intra prediction direction and the horizontal line. Fig. 1 shows both the case when the edges are aligned with the intra prediction direction and the case when they are not aligned.

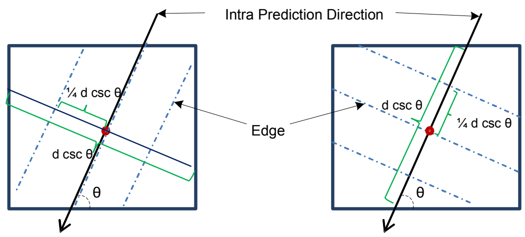


Fig. 1: Dependency between prediction direction and the edge modes

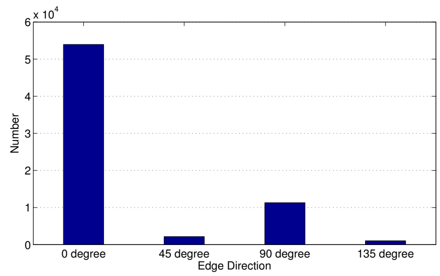


Fig. 2: Histogram of edge direction when intra prediction is horizontal in *SlideEditing*. The angle classifications are defined on the left side of Fig. 3

Computed over the first 50 frames of *SlideEditing*, Fig. 2 illustrates the histograms of four typical edge directions selected by the encoder during the RD-optimization process, when intra prediction is horizontal. As expected, the edge direction aligned with the prediction direction is used most frequently. The histogram also shows that the edge orientations are sometimes not aligned with the prediction direction. To cover these possibilities, three edge positions that are orthogonal to the intra predictions are also checked as illustrated in Fig. 1.

Thus, depending on the direction of a specific intra prediction mode, six edge positions are considered and the best one is selected by minimizing the RD cost

$$J = D_p + \lambda R_p, \tag{1}$$

where D_p and R_p are the corresponding distortion and number of bits associated with coding a block using an edge position denoted by p.

2.2. Simplification using Mode Classification

In order to achieve more accurate spatial prediction, HEVC supports a total of 35 intra prediction modes. The mode numbers and the corresponding prediction methods are shown in Fig. 3, where mode 0 is DC prediction, mode 1 is planar intra prediction, and modes 2 to 34 are directional modes covering 33 different prediction angles. Using six edge positions for each of the 33 directional prediction modes would require the encoder to perform up to almost 200 additional RD tests for each prediction block, which is impractical. To simplify the implementation of the proposed EM scheme, we classify intra modes into four main directions as described below.

Intra modes with an approximately horizontal prediction direction, modes 6-14, are classified to a group called *Degree 0*. Similarly, modes 22-30 are classified as *Degree 90*; modes 2-5 and 31-34 are classified as *Degree 45*; and modes 15-21 are classified as *Degree 135*, as shown in Fig. 3. With the above simplification, the number

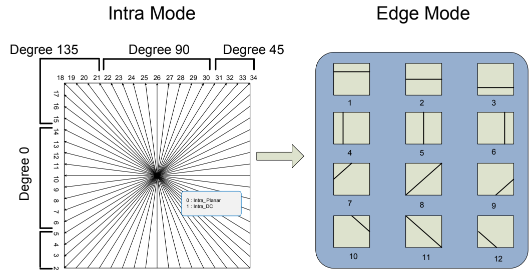
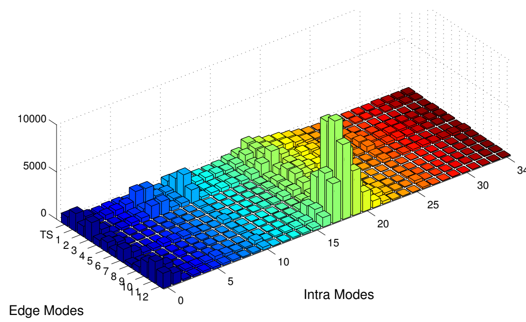
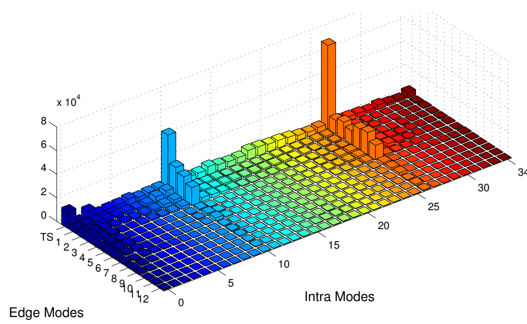


Fig. 3: Intra modes classification and edge modes



(a) BasketballDrillText



(b) SlideEditing

Fig. 4: Histogram of edge mode occurrence for different intra prediction modes

of allowed edge positions is restricted to 12 edge modes, indexed from 1 to 12 as shown in Fig. 3. For the intra DC mode and the intra planar mode, edge modes 1 to 6 are more likely to be used based on our observations, so they are classified to *Degree 0*.

To verify the dependency between intra prediction modes and edge modes, we ran encoding experiments to check the optimal edge mode for each intra prediction mode using the RD optimization process. In the experiments, 50 frames of various screen content videos are encoded using Intra mode The histograms of the best edge modes for two representative sequences, *BasketballDrillText* and *SildeEditing*, are shown in Figs. 4(a) and (b), respectively. As shown in Fig. 4a, we see that blocks associated with diagonally-oriented prediction modes (e.g., modes 16-20) tend to be coded using diagonal edge modes. In contrast, diagonal edge modes are less likely to be chosen for vertical and horizontal intra prediction modes. The same observation applies to Fig. 4b. Here we see that many vertical and horizontal edge modes are used with the vertical and horizontal prediction modes. These observations confirm that when vertical/horizontal intra predictions are used, vertical/horizontal edge modes are best suited for coding the residuals. When diagonal intra

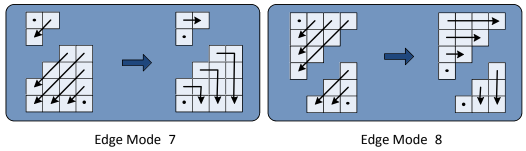


Fig. 5: 2D DCT transforms for diagonal edge modes

predictions are used, diagonal edge modes are more appropriate.

2.3. Applying Transforms to Sub-blocks

For each block that uses an edge mode, we partition it into two sub-blocks and then apply separable 2D DCT transforms. For edge modes 1-6, as pictured in Fig. 3, each sub-block is an $M \times N$ rectangle, so the existing horizontal and vertical transforms from HEVC can be applied. For edge modes 7-12, which partition the block into non-rectangular regions, several options similar to the shape-adaptive transform [19] can be considered. In this work, we develop directional 2D transforms that apply the DCT along separable paths in each partition

The directional 2D DCT used here comprises the following two steps:

1. First, a set of 1D DCTs is applied diagonally in alignment with the edge orientation. For example, as shown in Fig. 5, sub-blocks for edge modes 7 and 8 are first transformed along a diagonal direction. The DCT coefficients are also arranged from low to high frequencies along these paths. As a result, the lowest-frequency DCT coefficients are located along the top row and right column.
2. Next, the second set of 1D DCTs is applied along paths so that the DC coefficients of the first transform are covered by one DCT. The transform path for the AC coefficients are similar, as shown in Fig. 5.

2.4. Integration and mode coding in HEVC

2.4.1. Integration into HEVC

The proposed edge mode coding scheme is effective for blocks containing strong edges. For smooth blocks, the existing HEVC modes work best. Because screen content video often contains mixed natural and graphics material, we adaptively select among the existing and new edge modes for coding each intra prediction block. This combined scheme is called HEVC-plus-edge-mode (HEVC/EM). The encoder is illustrated in Fig. 6.

Up to three modes are checked for each intra block. As described earlier, a subset of edge modes are evaluated during RD-optimization, depending upon the intra prediction mode. Additionally, the unmodified HEVC transform is tested on the block, and transform-skip mode is checked as well, if enabled.

2.4.2. Coding of edge mode index

Recall that the edge modes are separated into a horizontal/vertical set and a diagonal set, where the intra prediction mode determines which set is used. Therefore, we only need signal one of six edge modes. We also need to signal whether the existing HEVC transform or transform skip mode are applied. Three bits are therefore needed to signal which of these eight modes to use. Table 1 lists the codewords for each mode. If the first two bits are zero, then the least significant bit is *0* when the existing transform from HEVC

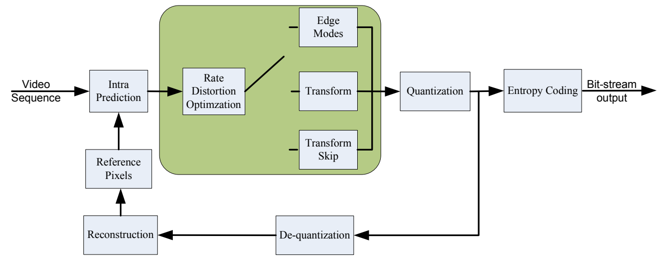


Fig. 6: Flowchart of the proposed HEVC/EM

Table 1: Codewords for edge modes

Code	Edge Mode	Code	Edge Mode
000	HM	100	3 or 9
001	TS	101	4 or 10
010	1 or 7	110	5 or 11
011	2 or 8	111	6 or 12

should be applied (denoted HM), and *1* indicates that the transform-skip mode (TS) from HEVC should be used. The remaining values are used to signal which edge mode to use from the subset of six possible modes.

3. EXPERIMENTAL RESULTS

The proposed HEVC/EM scheme was implemented using the HEVC test model reference software HM 7.0 [21]. It was tested using the intra main common test conditions [22]. The proposed HEVC/EM algorithm was tested against a reference using unmodified HM 7.0, for the four screen content sequences: *BasketballDrillText*, *ChinaSpeed*, *SlideEditing* and *SlideShow*. Note that transform skip (TS) [20] is not enabled for the intra main common test conditions. For comparison, we also present coding performance results for when TS alone is enabled. When enabled, TS or edge modes can be applied to 4x4 blocks.

For a typical encoding run, Fig. 7 shows where the edge modes and TS modes are used on one frame of *SlideEditing*. Fig. 7(a) shows a portion of the original picture. First, only TS mode is enabled.Blocks encoded with TS are labeled in red in Fig. 7(b). We can see that TS mode is used over most text areas, and the conventional HM transform was used over the smooth areas. This behavior is consistent with the expectation that the energy in blocks containing very sharp transitions is spread over many transform coefficients, thus reducing the coding efficiency of the transform as compared to using no transform at all.

Fig. 7(c) shows mode usage results for when both TS and edge modes are enabled. Blocks using TS are marked in red, and edge-mode blocks are marked in blue. We observe that both the edge mode and TS are applied in areas containing strong edges. Moreover, many blocks marked in red in Fig. 7(b) are now covered by blue, which indicates that the edge modes are more efficient, in a rate-distortion sense, than TS for coding these areas.

To evaluate the RD performance, four test sequences (Class F) were coded using QP values of 22, 27, 32, and 37, respectively. 500 frames were coded for sequences *BasketballDrillText*, *ChinaSpeed*, and *SlideShow* and 300 frames were coded for sequence *SlideEditing*. The performance of the original HM is used as the reference, and BD-rate [23] changes are calculated for when TS alone is enabled and for when the EM modes are additionally included. For informational purposes, we also show results denoted by EMD, for

(a) Original

(b) TS

(c) TS and EM

Fig. 7: Location of blocks encoded with TS (red) or edge (blue) modes

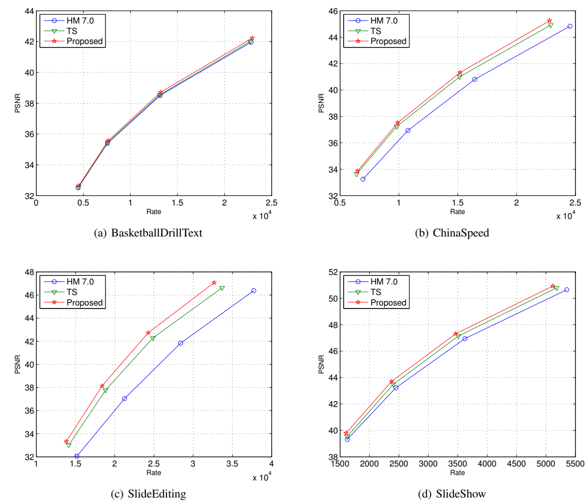
when EM modes are available but TS is disabled by not allowing codeword 001 from Table 1 to be used. In practice, further improvement for the EMD case would be possible by redesigning the codewords so as not to leave 001 unused.

The experimental results are summarized in Table 2. A negative change in BD-rate indicates improvement in compression efficiency of the tested method over the reference. TS achieves an average bit-rate reduction of 7.5%, and incorporating EM yields a 10.4% reduction. Disabling TS while keeping edge modes yields 8.2%, indicating that the improvements from edge modes are additive. Among the four screen content test sequences, less improvement is achieved in *BasketballDrillText*. This is reasonable given that the majority of its content is natural, while only a strip of graphics material is embedded in the sequence. Including edge modes improves upon TS for that sequence, especially in areas with diagonal edges such as the basketball net. RD curves are shown in Fig. 8, showing that the proposed HEVC/EM scheme offers the best RD performance.

For completeness, we also investigated the performance of HEVC/EM and TS on natural video sequences, which include the class A through class E material listed in the common test conditions [22] Experimental results are summarized in Table 3. For classes A, B, and E, there is little or no change in RD performance using edge modes, while there is slight improvement for the lower-resolution Class C and D sequences. Although the proposed HEVC/EM require additional bits to signal the modes, its coding performance for natural sequences does not decrease. Transform skip has little to no impact on performance for all classes.

Table 2: BD-Rate changes for screen content sequences using transform skip (TS) alone, incorporating edge modes (EM), or edge modes with TS disabled (EMD)

	△ BD-Rate (%)		
Sequence	**TS**	**EMD**	**EM**
BasketballDrillText	-0.7	-2.5	-2.9
ChinaSpeed	-10.5	-9.1	-13.0
SlideEditing	-14.5	-14.7	-18.0
SlideShow	-4.4	-6.6	-7.7
Average	**-7.5**	**-8.2**	**-10.4**



(a) BasketballDrillText

(b) ChinaSpeed

(c) SlideEditing

(d) SlideShow

Fig. 8: Comparison of RD curves for HM, TS and HEVC/EM

4. CONCLUSIONS

A new coding tool using edge modes, HEVC/EM, was introduced and integrated into the HEVC test model (HM). Several edge orientations and positions based on intra prediction directions were tested on screen content material to show that performance could be improved by partitioning a prediction block into two sub-blocks. After partitioning, a rectangular 2D DCT or a directional 2D DCT is applied to these sub-blocks. Experiments show that the proposed HEVC/EM scheme is effective in coding blocks with strong edges, yielding up to an 18% reduction in bit-rate and an average reduction of 10.4% for screen content video. HEVC/EM offers a significant coding gain over unmodified HM, and it provides additional improvement over the existing HEVC transform skip mode when coding screen content video.

Table 3: BD-Rate change for different classes of natural sequences

	△ BD-Rate (%)	
Sequence set	**TS**	**EM**
Class A	0.0	0.0
Class B	0.0	-0.1
Class C	0.0	-0.8
Class D	0.0	-1.0
Class E	0.1	0.0
Average	**0.0**	**-0.4**

5. REFERENCES

[1] Y. Lu, S. Li and H. Shen, "Virtualized screen: a third element for cloud–mobile convergence," *IEEE Multimedia*, Vol. 18, no. 2, pp. 4-11, Apr. 2011.

[2] C. Lan, X. Peng, J. Xu and F. Wu, "Intra and inter coding tools for screen contents," *Document of Joint Collaborative Team on Video Coding*, JCTVC-E145, Mar. 2011.

[3] S. Wang and T. Lin,"A Unified LZ and hybrid coding for compound image partial-lossless compression," in proc. *IEEE Int. Congress on Image and Signal Processing*, pp. 1-5, Oct. 2009.

[4] W. Ding, Y. Lu and F. Wu, "Enable efficient compound image compression in H.264/AVC intra coding," *IEEE Int. Conf. on Image Processing*, pp.337-340, Sep. 2007.

[5] K. Konstantinides and D.Tretter, "A JPEG variable quantization method for compound documents," *IEEE Trans. on Image Process.*, Vol. 9, no. 7, pp. 1282-1287, Jul. 2000.

[6] Z. Pan, H. Shen, Y. Lu, and S. Li, "Browser-friendly hybrid codec for compound image compression," *IEEE International Symposium on Circuits and Systems*, pp. 101-104, May 2011.

[7] T. Lin, P. Zhang, S. Wang, K. Zhou and X. Chen, "Mixed chroma sampling-rate High Efficiency Video Coding for full-chroma screen content," *IEEE Trans. Circ. Syst. Video Technol.*, to appear.

[8] A. Zaghetto and R. L. de Queiroz, "Segmentation-driven compound document coding based on H.264/AVC-INTRA," *IEEE Trans. on Image Process.*, vol. 16, pp. 1755-1760, Jul. 2007.

[9] T. Wiegand, W. Han, B. Bross, J. Ohm, and G. J. Sullivan, "WD5: Working draft 5 of High-Efficiency Video Coding," *Document of Joint Collaborative Team on Video Coding*, JCTVC-G1103, Geneva, CH, Nov. 2011.

[10] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560-576, Jul. 2003.

[11] T. Lin and P. Hao, "Compound image compression for real-time computer screen image transmission," *IEEE Trans. Image Process.*, vol.14, no. 7, pp. 993-1005, Jul. 2005.

[12] L. Bottou, P. Haffner, P. Howard, P. Simard, Y. Bengio, and Y. LeCun, "High quality document image compression with DjVu," *J. Electron. Imag.*, vol. 7, no. 3, pp. 410-425, Jul. 1998.

[13] R. L. de Queiroz, Z. Fan, and T. Tran, "Optimizing block-thresholding segmentation for multilayer compression of compound images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 1461-1471, Oct. 2000.

[14] H. Cheng and C. A. Bouman, "Document compression using rate-distortion optimized segmentation,"*J. Electron. Imag.*, vol. 10, pp. 460-474, Apr. 2001.

[15] C. Lan, G. Shi and F. Wu, "Compress compound images in H.264/MPEG-4 AVC by exploiting spatial correlation," *IEEE Trans. on Image Process.*, vol. 19, no. 4, pp. 946-957, Apr. 2010.

[16] M. Mrak and J.-Z. Xu, "Improving screen content coding in HEVC by transform skipping," *EUSIPCO 2012*, pp. 1209-1213, Aug. 2012.

[17] J. Nam, D. Sim, and I. V. Bajic, "HEVC-based adaptive quantization for screen content videos," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting,* pp. 1-4, June 2012.

[18] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. on Image Process.*, vol. 11, no. 6, pp. 670–684, June 2002.

[19] T. Sikora and B. Makai, "Shape-adaptive DCT for generic coding of video," *IEEE Circ. Sys. Video Technol.*, vol. 5, no. 1, pp. 59-62, Feb. 1995.

[20] C. Lan, X. Peng, J. Xu and F. Wu, "Intra transform skipping," *Document of Joint Collaborative Team on Video Coding*, JCTVC-I0408, Apr. 2012.

[21] HM 7.0 software, available online `https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/`

[22] F. Bossen, "Common test conditions and software reference configurations," *Document of Joint Collaborative Team on Video Coding*, JCTVC-H1100, San Jose, CA, Feb. 2012.

[23] G. Bjontegaard, "Calculation of average PSNR difference between RD-curves", VCEG-M033, 13th VCEG meeting, Austin, TX, Apr. 2001.