TWO HEVC ENCODER METHODS FOR BLOCK ARTIFACT REDUCTION

Andrey Norkin, Kenneth Andersson, and Valentin Kulyk
Ericsson Research, Ericsson, Stockholm, Sweden

ABSTRACT

The HEVC deblocking filter significantly improves the subjective quality of coded video sequences at lower bitrates. During the final phase of HEVC standardization, it was shown that the reference software encoder may produce visible block artifacts on some sequences with content that shows chaotic motion, such as water or fire. The paper analyses the reasons for blocking artifacts in such sequences and describes two simple encoder-side methods that improve the subjective quality on these sequences without degrading the quality on other content and without significant bitrate increase. The effect on subjective quality has been evaluated by a formal subjective test.

Index Terms— Deblocking, HEVC, subjective quality, block artifacts, encoder

1. INTRODUCTION

The HEVC standard has finally been finalized in January 2013 [1], [2]. The standard substantially improves compression of video, especially on the high resolution content. The new standard is different in many aspects from the H.264/AVC video coding standard. For most of the previous MPEG-x and H.26x codecs, the largest independently encoded entity was a 16 × 16 pixels macroblock. For HEVC, the picture is split into coding-tree units (CTU) with a maximum size of 64 × 64 pixels. Every CTU is a root of a quadtree, which can be further divided into leaf-level coding units (CU) in a quadtree fashion. Each CU contains one or more prediction units (PU) that are predicted independently of each other. A CU is also associated with a transform quadtree that compresses the prediction residual and has a structure similar to that of a CTU. The leaves of the quadtree are transform units (TU). Partitions for motion prediction have square or rectangular shapes. The size of the prediction blocks in HEVC can vary from 4 × 4 samples to 64 × 64, while transform sizes vary from 4 × 4 to 32 × 32 samples. The size of intra-predicted partitions varies from 4 × 4 to 32 × 32 luma samples.

HEVC uses two in-loop filters: a deblocking filter and a sample adaptive offset (SAO) filter, which is applied to the output of the deblocking filter. In-loop filters improve the subjective quality of reconstructed video as well as compression efficiency.

HEVC deblocking filter is intended to attenuate artifacts that appear due to relatively independent encoding of blocks in a picture. However, despite of the fact the HEVC deblocking improves the subjective quality and compression efficiency, some blocking artifacts have been reported for lower bitrates on the sequences with high level of chaotic motion [11]. This problem has been brought up during the late phase of the HEVC standardization [4], [5], [11]. This paper provides analysis of the problem and suggests two simple encoder side methods that are capable of improving the quality on such sequences while still keeping good subjective quality on the "normal" video content.

The paper is organized as follows. First, a short description of the HEVC deblocking filter is given in Section 2. Section 3 provides analysis of the reason for block artifacts in content with chaotic motion. Sections 4 and 5 describe two methods that address the problem of the block artifacts in such content. The test setup and test methodology used to evaluate the subjective quality are described in Section 6 while Section 7 presents the results. Finally, Section 8 concludes the paper.

2. HEVC DEBLOCKING FILTER DESCRIPTION

A good deblocking filter needs accurate decisions whether the block edge is to be filtered or not since excessive smoothing of the highly detailed area may decrease the subjective quality. Filtering criteria are used to determine whether discontinuities at the block boundary are likely due to prediction or quantization error or they are a part of the original sequence. To make this decision, the video coding parameters are used along with characteristics of the signal at the sides of the block boundary.

In HEVC deblocking filter [3], only the block boundaries that lie on the 8 × 8 pixel grid are filtered in order to decrease the worst case complexity. The deblocking filter is only applied to the block boundaries that correspond to the CU, PU or TU boundaries if at least one of the following conditions is fulfilled: at least one of the adjacent blocks is intra-predicted or has non-zero transform coefficients, the difference between the motion vectors of the adjacent blocks is at least one integer sample or motion vectors point to different reference frames [3].

Additionally, for each block boundary of four samples in length that satisfies the conditions above, the signal structure on both side of the block boundary is evaluated to decide whether the deblocking filtering is applied, the strong or normal deblocking filter is used and how many samples are to be filtered [3]. The following expression is used to evaluate whether the deblocking is applied to the block boundary.

$$| p_{2,0} - 2p_{1,0} + p_{0,0} | + | p_{2,3} - 2p_{1,3} + p_{0,3} | + \\ | q_{2,0} - 2q_{1,0} + q_{0,0} | + | q_{2,3} - 2q_{1,3} + q_{0,3} | < \beta, \quad (1)$$

where threshold β depends on the quantization parameter QP that is used to adjust the quantization step for prediction error coefficients, $p_{i,j}$ is the pixel value on row j and i samples away from the block boundary on the left side of the boundary and $q_{i,j}$ is corresponding sample on to the right from the boundary. The type of deblocking filtering and the number of modified samples also depend on parameter β. For the exact expressions used to evaluate these parameters the readers are referred to [2] and [3]. When the normal deblocking filtering is applied, the pixels are modified as

$$p_i' = p_i + \Delta_{pi}. \quad (2)$$

where p_i and p_i' are the values of the sample before and after modification respectively, and Δ_{pi} is obtained by filtering followed by a clipping operation. The clipping operation is therefore used to limit the degree of filtering in order to avoid excessive smoothing. The clipping is applied by performing the following operation:

$$\Delta = \mathrm{Min}\,(\mathrm{Max}(-c,\ \delta\,),\, c), \quad (3)$$

where the value of c is equal to t_C for p_0 and q_0, and $t_C/2$ for p_1 and q_1 in the case of normal filtering. In the case of strong filtering, c is set equal to $2t_C$. Variable t_C is obtained from a table as t_C(QP) when both adjacent blocks are inter-predicted and t_C(QP+2) when one of adjacent blocks is intra-predicted [2], [3].

The dependency of parameters β and t_C on QP is shown in Figs. 1 and 2. One can observe that the values of β and t_C increase with QP. Therefore, the deblocking is applied more frequently at high QP values compared to low QP values and larger modifications of sample values by deblocking filtering are allowed. The deblocking operation is effectively disabled for low QP values by setting one or both of β and t_C to zero.

The deblocking parameters t_C and β provide adaptivity to the QP and prediction type. The deblocking can be further adjusted on a slice or picture level by sending parameters in the slice header or picture parameters set (PPS) to control the amount of filtering. The corresponding parameters are tc_offset_div2 and beta_offset_div2. The specify the offsets (divided by two) that are added to the QP value before determining the β and t_C. Roughly speaking, parameter beta_offset_div2 adjusts the number of pixels to which the deblocking is applied, whereas parameter tc_offset_div2 adjusts the amount of filtering that can be applied to those pixels, as well as detection of natural edges.

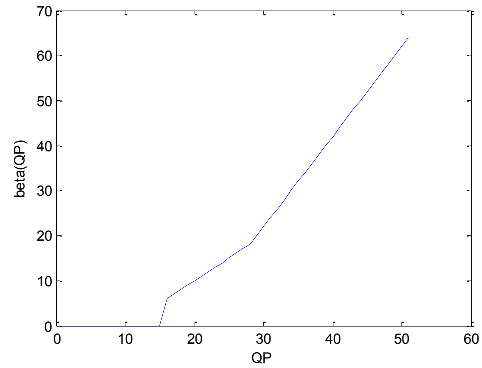


Fig . 1. Dependency of β on QP

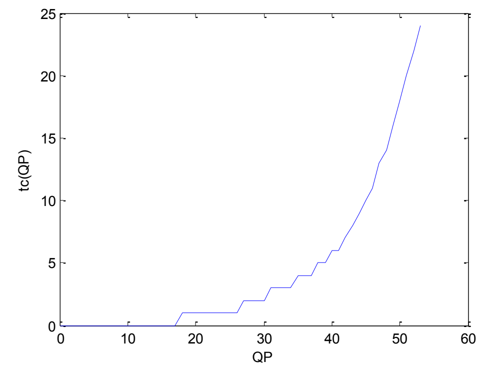


Fig. 2. Dependency of t_C on QP

3. BLOCK ARTIFACT ANALYSIS

As mentioned earlier, it has been reported at a later stage of the standardization process that some remaining block artifacts are present in the sequences that have content with much chaotic motion, such as water, rain, snow, fire etc. For example, one can see remaining block artifacts in Fig. 3

Several explanations can be found for these remaining artifacts. Both random-access and low-delay test configurations in HEVC toggle the picture quality in hierarchical manner depending on picture position. For example, random access configuration uses hierarchical-B GOP8 structure (see Fig. 4), where the base QP is used for the intra-coded pictures, QP+1 is used for the B-pictures at depth 0 (e.g. POC 8), i.e. pictures used at the lowest depth, QP + 2 for the pictures at coding depth 1. Depth 2 uses QP + 3 and non-reference b-pictures at depth 3 use QP + 4.

Fig. 3. Remaining block artifacts (sequence *Riverbed*, base QP 37)

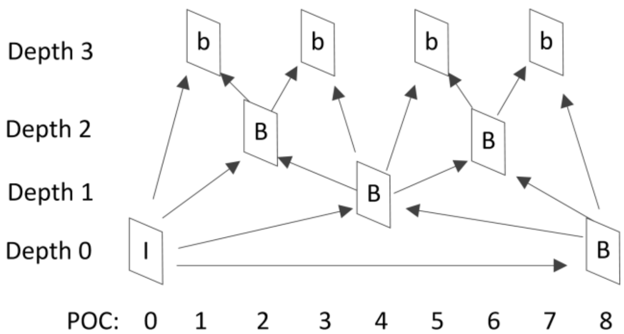


Fig. 4. Example of hierarchical-B coding structure with GOP8 illustrating different depth of the picture in the coding structure

In addition to toggling the base QP value for pictures coded at different depth, different λ parameter [10] is used in the mode decisions, motion estimation, and rate-distortion optimized quantization. Rate-distortion optimization in video codecs is often done with the Lagrange multipliers method, i.e. by finding the minimum cost J determined as

$$J = D + \lambda R, \tag{4}$$

where D is distortion, R is the rate (or a number of bits) and λ is the Lagrange multiplier. As one can see from (4), lower values of λ make the encoder choose coding modes that use higher rate and achieve lower distortion while higher values of λ favor the coding modes resulting in lower rate and higher distortion. The value of λ is usually linked to the QP. The value of λ can also depend on the picture depth. The example of parameter λ in the random access configuration of the HEVC common test conditions is shown in Fig. 5. One can see that λ is higher for the pictures with higher depth.

By inspection of encoded videos, it has been found that block artifacts mostly appear in the pictures with higher depth, in particular in the non-reference pictures at depth 3 and reference pictures at depth 2. Some milder artifacts are present in pictures at depth 1. Pictures at depth 0 usually have higher quality and do not exhibit block artifacts even at base QP equal to 37. This observation is aligned the fact that λ has different dependency on QP for pictures with different depth in coding hierarchy (see Fig. 5).

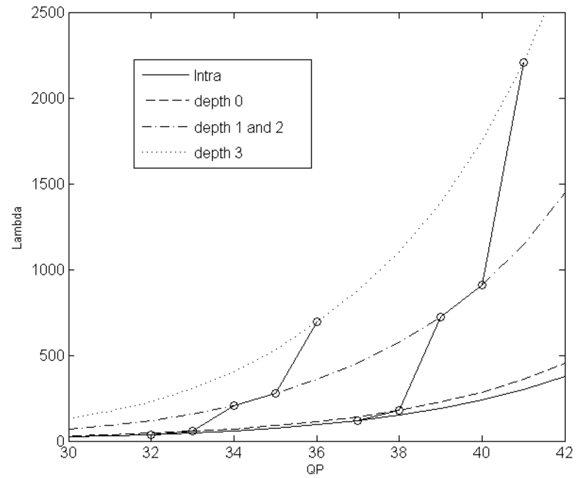


Fig. 5. Dependency of λ from QP for pictures at different depth. Values of λ for picture structures at base QP 32 and base QP 37 are also shown.

One can see in Fig. 5 that λ for pictures at depth 3 at base QP 32 (picture QP 36) is approximately the same as lambda parameter at base QP 37 and depth 1 (picture QP 39) and significantly higher than lambda for pictures with depth 0 at picture QP 38.

Another observation from studying the encoded sequences is that visible block artifacts are often associated with large intra-blocks (e.g. 32 × 32 blocks), although the artifacts are sometimes also present between inter-predicted blocks. When the motion is chaotic it is difficult to get good prediction from the previously decoded pictures. In such cases, intra-picture prediction is often chosen by the encoder.

Summarizing the written above, block artifacts generated by the HEVC reference software often occur in pictures using high values of parameter λ (i.e. pictures at high depth in coding hierarchy), in particular at large intra-predicted blocks in these pictures. The following two sections propose the methods that can be used to improve the subjective quality on difficult sequences without changing the lambda parameter and affecting the quality of the "normal" video sequences content.

4. HIERARCHICAL DEBLOCKING STRENGTH ADAPTATION

In order to attenuate block artifacts in a picture with higher depth, HEVC can be configured to signal the deblocking filter offsets at the slice/picture level. Higher offsets are sent for the frames, which are at higher depth in the coding hierarchy.

For this article, we have tested somewhat extreme values of deblocking offsets in order to study the effect of these offsets on "normal" sequences. The parameters used in the experiment are provided in Table 1.

Table 1. Deblocking parameters for pictures at different depth

Depth	tc_offset_div2	beta_offset_div2
Intra pic.	0	0
0	1	0
1	3	0
2	4	1
3	6	2

The value 6 is the value allowed by the standard for tc_offset_div2, which corresponds to a shift by 12 when choosing the tc_offset(QP) value from the table.

The proposed approach relaxes the deblocking decisions thresholds and clipping values for the pictures at higher depth, which results in deblocking being applied more frequently to these pictures, modifying larger number of samples from the block boundary and allowing larger modifications to the samples.

When it is desirable to limit the potential influence of the approach on other types of video content, the deblocking parameters can be changed more conservatively. For example, a tc_offset_div_2 can take values 0, 3, 3, and 5 for pictures at depths 0, 1, 2, and 3 respectively, while the beta_offset_div2 is set to zero for all the pictures.

5. REDUCING INTRA BLOCK SIZE

The HEVC deblocking filtering can only modify three pixels from the block boundary. When the rate-distortion optimization chooses 32×32 intra-predicted CUs at higher depth, the prediction is often coarse and since the transform coefficients are also coarsely quantized, block artifacts can appear. It might be difficult to conceal a blocking artifact by just applying the deblocking filtering. Imagine 32×32 lighter and darker uniform blocks arranged in a checkerboard pattern. Even if the deblocking is applied to all block boundaries and transitions between the blocks are smooth, the block pattern would still be visible since there are only three pixels from each side of the block boundary that are modified while most samples in the block are not changed by deblocking.

To improve visual quality, the HEVC reference encoder can limit the maximum TU size to 16×16 samples for coding of intra CUs in inter-predicted slices by setting the configuration parameter RDpenalty to 2. This constraint will also restrict the maximum intra-predicted block size and may increase the bitrate from 0 to a couple of percent. Alternatively, a penalty for using 32×32 intra-prediction blocks can be applied to the rate-distortion cost by setting the RDpenalty to 1. In our experiments, RDpenalty has been set equal to 2.

6. SUBJECTIVE TEST SETUP

6.1 Test method and environment

The subjective method Double Stimulus Continuous Quality Scale (DSCQS) variant I (one viewer at a time) [7] has been chosen for this subjective test since deblocking has more effects at medium and low video quality condition. This method is particularly useful when the test does not use full range of quality. Moreover, according to the research results, the DSCQS method is better than the other test methods in minimizing contextual effects for subjective quality assessment [7].

The test has been conducted with naïve viewers. It consisted of two sessions of 20 minutes each. A test subject was presented a series of video pairs in random order. The processed video sequences (PVS) were presented in pairs (“A” and “B”) and every pair consisted of the same video sequence where one test video was processed with the anchor (reference) and the second video was processed with the method under assessment. The subjects were asked to assess the overall quality of each video in the pair using the respective voting scale without knowing which one was the reference video. The video sequences could have been viewed several times (maximum three times). Both playout and voting procedure was steered by test subjects who could initiate the voting when ready for it. A training session was held after the instructions and before the test execution to help test subjects to understand the test procedure and familiarize them with typical video quality conditions. The training session comprised six pairs of video sequences, different from those used in the test.

The subjects scored the videos on a continuous quality assessment scale (maximum value: 100, minimum value: 0) based on five categories “Excellent”, “Good”, “Fair”, “Poor” and “Bad” (in Swedish).

Each test session involved only one subject per display assessing the test material. The subject was placed directly at the center of the display at the 3H (three times of the display height) viewing distance. The ambient lighting system based on high frequency neon lamps with color temperature of 6500 K was used.

The test environment setup and display settings are summarized below:

- Display: 40” Sony KDL-40X3500
- Display settings: 50Hz refresh rate, Brightness – 48 (~200cd/m2 peak luminance), Contrast - 90, Backlight – 6, Sharpness – 60, Noise Reduction – off.
- The viewing distance was 3 times the display height.

- The ambient light level measured at viewing position was ~25 lux.

In total, 38 naïve viewers, not working professionally with quality assessment or video processing participated in the test. All viewers were screened for normal visual acuity (20/30 for one eye and 20/25 for both eyes) using Snellen test with or without corrective glasses. All subjects had visual acuity 20/25 or higher.

6.2 Test material

In total, eight 1080p25 source video sequences with duration 10 seconds were used in the test. Three sequences were taken from the JCT-VC test set (Riverbed, ParkScene, Kimono) and five other sequences were prepared for the test (Fire, Fountain, Water1, Water6 and Underwater). All source sequences contained high amount of motion except ParkScene and Kimono. These two videos have been used to verify that the proposed methods don't impair the video quality compared to the anchor. ParkScene and Kimono sequences originated from 1080p24 source were played at 25 Hz frame rate to avoid impact of the playback jitter on the quality.

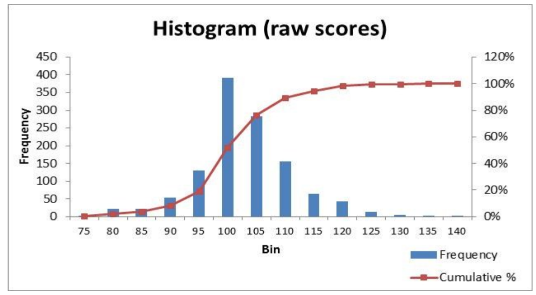
Both the anchor and proposed approach described in Sections 4 and 5 were coded with HM-11.0 [6] in random access configuration.

6.2 Subjective data evaluation

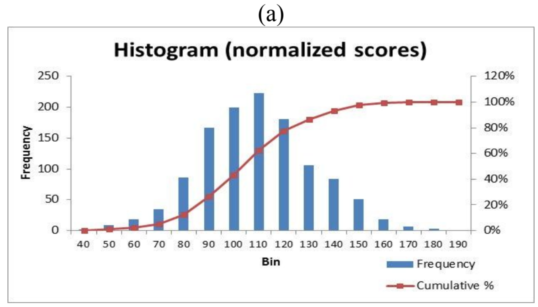
The two-way ANOVA was applied to the raw difference scores to check the between-subject variability and that scores given by subjects were drawn from the same distribution (under the null hypothesis for between-subjects variation). The results of the ANOVA indicated that there was significant ($p = 0$) subject-to-subject variability. Therefore the scores were also normalized according to offset mean correction rule [8], [9]. The results for both raw and normalized scores were analyzed. The two-way ANOVA applied after the mean offset correction showed that between-subject variation decreased ($p = 0.19 > 0.05$).

The screening of possible outlier subjects was performed on both data sets, according to the procedure described in [7]. Four of 38 subjects were detected and discarded from both the normalized and raw results. The results of the normalization procedure can be observed by comparing the histograms (see Fig.6).

After the screening, the mean opinion scores were calculated for every PVS. The difference between the reference video score and the accessed video score was subtracted from the maximum value of the voting scale (100): *100 – (ref_score – assessment_score)*. Then the difference values were averaged across all the subjects to yield a difference mean opinion score (DMOS) for non-normalized and normalized subjective data (NDMOS).



(a)



(b)

Fig. 6. Histograms showing distribution of scores: (a) raw scores, (b) normalized scores.

Thus, DMOS and NDMOS scores below 100 correspond to cases when the test subject considered the reference video quality higher than the assessed video and above 100 when the assessed video (from the proposed method) was judged to have higher quality than the reference.

7. RESULTS

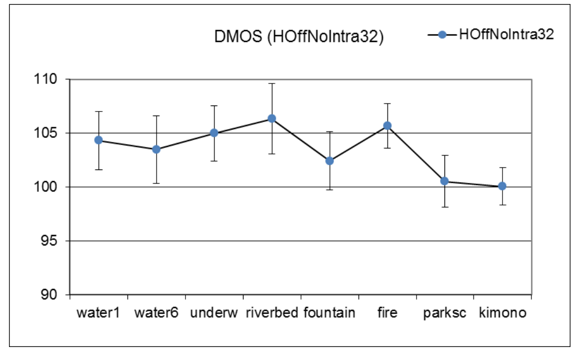
Fig. 8 shows test results per video sequence, both DMOS and normalized DMOS (NDMOS). One can see that the results for both DMOS and NDMOS are quite similar. All six difficult sequences were scored higher than the reference with non-overlapping confidence interval for the normalized scores and five out of these six sequences show non-overlapping confidence intervals with the reference for the raw scores. Once can also see that the two sequences used to check that the proposed method does not impair the normal sequences show very similar results to the reference.

The bitrates of the sequences generated by the proposed method are shown in Table 2. One can see that the bitrates are on average similar to the bitrates of the anchor with the biggest difference of 4.9% on the Underwater sequence. The difference is caused by not choosing 32 × 32 intra-predicted blocks in reference frames. On the sequence Fire, however, the bitrate difference is below 1%, while this sequence exhibits the second biggest subjective quality difference between the tested method and the reference. Fig. 7 demonstrates how the proposed approach improves visual quality (compared with the reference shown in Fig. 3).

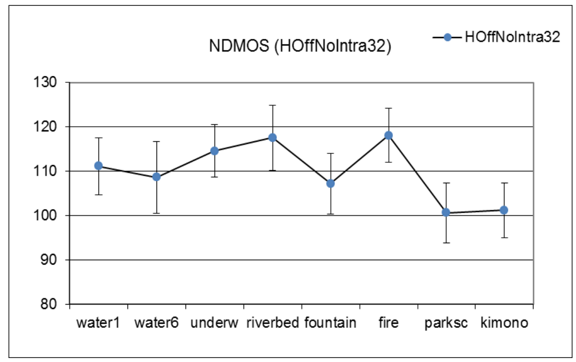
Table 2. Bitrates of the proposed methods compared to the anchor

Video	Anchor		HOffNoIntra32		
	QP	[kbps]	QP	[kbps]	%
Fountain	37	6104.972	37	6157.793	100.9
Fire	37	2441.594	37	2462.629	100.9
Riverbed	37	2346.436	37	2440.269	104.0
Park Scene	37	717.301	37	718.150	100.1
Underwater	37	666.062	37	698.535	104.9
Water6	37	616.439	37	628.155	101.9
Kimono	37	542.574	37	544.578	100.4
Water1	32	541.010	32	559.620	103.4

Fig. 7. Proposed encoder approaches – compared with reference in Fig. 3 (sequence *Riverbed*, base QP 37).



(a) Raw scores



(b) Normalized scores

Fig. 8. Test results based on (a) raw scores; (b) normalized scores.

8. CONCLUSIONS

The obtained results indicate that the proposed methods improve the subjective quality on difficult sequences, such as the sequences that contain water, rain, fire etc. The proposed methods do not impair the quality of the video sequences that have slow or linear motion, i.e. "normal" video content. These two methods offer a practical solution that improves the subjective quality significantly when encoding the sequences with chaotic motion with HEVC encoder.

9. REFERENCES

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

[2] Recommendation ITU-T H.265, “High efficiency video coding”, April 2013.

[3] A. Norkin, G. Bjøntegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson, M. Zhou, and G. Van der Auwera, "HEVC deblocking filter", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1746 – 1754, Dec. 2012.

[4] A. Norkin, “Non-CE1: Non-normative improvement to deblocking filtering”, JCTVC-K0289, Shanghai, China, 10–19 Oct. 2012.

[5] A. Norkin, K. Andersson, R. Sjöberg, “AHG6: On deblocking filter and parameters signaling”, JCTVC-L0232, Geneva, Switzerland, 14–23 Jan. 2013.

[6] HEVC reference software, https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-11.0

[7] Recommendation ITU-R BT.500-13, “Methodology for subjective assessment of the quality of the television pictures”, 2012.

[8] E. D. Gelasca, “Full-reference objective quality metrics for video watermarking, video segmentation and 3D model watermarking”, Ph.D. thesis, EPFL, September 2005.

[9] F. D. Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro, and T. Ebrahimi, “Subjective quality assessment of H.264/AVC video streaming with packet losses”, *EURASIP Journal on Image and Video Processing*, vol. 2011, 12 pages.

[10] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression", *IEEE Signal Processing Magazine*, pp. 74 – 90, Nov. 1998.

[11] D.-K. Kwon, M. Budagavi, "Transform size dependent deblocking filter", JCTVC-I0244, Geneva, CH, Apr. – May 2012.