





biblio.ugent.be

The UGent Institutional Repository is the electronic archiving and dissemination platform for all UGent research publications. Ghent University has implemented a mandate stipulating that all academic publications of UGent researchers should be deposited and archived in this repository. Except for items where current copyright restrictions apply, these papers are available in Open Access.

This item is the archived peer-reviewed author-version of:

No-Reference Bitstream-Based Impairment Detection For High Efficiency Video Coding

Glenn Van Wallendael, Nicolas Staelens, Lucjan Janowski, Jan De Cock, Piet Demeester, Rik Van de Walle

Proceedings of Workshop on Quality of Multimedia Experience, Yara Valley, Australia, July 2012.

To refer to or to cite this work, please use the citation to the published version:

G. Van Wallendael, N. Staelens, L. Janowski, J. De Cock, P. Demeester, R. Van de Walle, "No-Reference Bitstream-Based Impairment Detection For High Efficiency Video Coding", Proceedings of Workshop on Quality of Multimedia Experience, Yara Valley, Australia, July 2012

NO-REFERENCE BITSTREAM-BASED IMPAIRMENT DETECTION FOR HIGH EFFICIENCY VIDEO CODING

Glenn Van Wallendael^a, Nicolas Staelens^b, Lucjan Janowski^c, Jan De Cock^a, Piet Demeester^b, Rik Van de Walle^a

^aGhent University - IBBT, Department of Electronics and Information Systems, Ghent, Belgium
^bGhent University - IBBT, Department of Information Technology, Ghent, Belgium
^cAGH University of Science and Technology, Poland

ABSTRACT

Video distribution over error-prone Internet Protocol (IP) networks results in visual impairments on the received video streams. Objective impairment detection algorithms are crucial for maintaining a high Quality of Experience (QoE) as provided with IPTV distribution. There is a lot of research invested in H.264/AVC impairment detection models and questions rise if these turn obsolete with a transition to the successor of H.264/AVC, called High Efficiency Video Coding (HEVC). In this paper, first we show that impairments on HEVC compressed sequences are more visible compaired to H.264/AVC encoded sequences. We also show that an impairment detection model designed for H.264/AVC could be reused on HEVC, but that caution is advised. A more accurate model taking into account content classification needed slight modification to remain applicable for HEVC compression video content.

Index Terms— Quality of Experience (QoE), Objective video quality, High Efficiency Video Coding (HEVC)

1. INTRODUCTION

With Internet Protocol Television (IPTV), video is distributed over an IP-based network which is prone to packet loss. A best-effort network introduces errors in the video stream resulting in visual distortions in the decoded video. Depending on the strength of the distortion introduced on the video stream, people will notice or even be annoyed by the resulting deformation. Research indicates that only one noticeable artefact per hour will be tolerated by end users watching the video [1]. For this reason, knowing the quality of the delivered video stream is valuable information for video service providers. This information enables them to monitor the Quality of Experience (QoE) of the delivered service and thus provides an indication for when improvements need to be considered.

When monitoring the quality of a video stream, the best indication a video service provider can obtain is the subjective opinion of the end users. This subjective opinion is indicated by means of a Mean Opinion Score (MOS). This is the score with which end users rate the subjective quality of a distorted video stream.

Because of the infeasibility of large scale subjective quality evaluation, objective quality metrics are being developed. These objective metrics try to provide similar information as end users would have provided with a subjective experiment. Different metrics exist [2], but in general a classification is made based on the amount of information used from the original video stream and the type of information used from the received video. According to the first classification of objective metrics, three different types can be differentiated, namely full-reference [3], reduced-reference [4], and no-reference metrics [5]. With the full-reference metric, the original video data is entirely used to derive a quality indication from. The best known full-reference metric used for video material is called Peak Signal to Noise Ratio (PSNR). When only a reduced amount of information from the original video stream can be used, the objective metric is classified as a reducedreference metric. Finally, when no access is provided to the original video stream, the metric is called no-reference metric. In an IPTV environment, only the reduced-reference and no-reference metric are deemed to be practical because the original video sequence is not accessible on the end-user devices where the quality is measured.

When classifying objective metrics on the type of information used from the received video, three different classes can be distinguished, namely parametric-based metrics [6], bitstream-based [5], or pixel-based. To explain the difference between these types, the layers traversed by the video stream need to be explained. On the lowest layer, video is transmitted over the IP-based network. With information from the IP network, it is possible to get an indication of the quality of the transported video stream. The metrics using network information are called parametric quality metrics. In a second phase, the video stream is reconstructed and elements like the prediction type of the video picture and the motion information contained in the video picture can be used. When video stream information is used to derive the video quality, the metric is called a bitstream-based metric. Next, the video stream is decoded and pixels are presented to the end user. When algorithms use displayed pixels to formulate a quality indication, the metric is a pixel-based metric. Finally, the three different levels of information gathering can be combined resulting in a hybrid quality metric [7].

Until now, a lot of research is invested in quality metrics for video streams compressed with H.264/AVC [8]. At this moment, a successor of this nine year old standard is being developed within a collaboration of the ISO/IEC Moving Picture Experts Group (MPEG) and ITU-T Video Coding Experts Group (VCEG) called High Efficiency Video Coding (HEVC) [9]. In this paper, the consequences of network impairments on HEVC based IPTV distribution will be compared to H.264/AVC based distribution. Additionally, a noreference bitstream-based video impairment detection developed for H.264/AVC [10] will be tested on its applicability for HEVC based video streams. In order to explain the impact of a switch from H.264/AVC to HEVC, relevant differences and resemblances between H.264/AVC and HEVC will be explained in Section 2. Next, the methodology used to evaluate the difference between packet loss effects on H.264/AVC and HEVC will be explained in Section 3. Results from the subjective evaluation of packet loss effects on HEVC and from the applicability of an H.264/AVC model on HEVC will be elaborated on in Section 4. Finally, a conclusion is provided in Section 5.

2. HIGH EFFICIENCY VIDEO CODING

Similar to the standardization process of H.264/AVC, a collaboration between MPEG and VCEG called Joint Collaborative Team on Video Coding (JCT-VC) is constructed to lead the standardization process of HEVC. This video compression standard under development is planned to be finalized by January 2013, but at this moment the standard is in a stable phase, making it suitable for analysis in different video application domains.

Like H.264/AVC, HEVC is still a block-based hybrid video compression standard. In a video compression context, blockbased means that the image is still divided in blocks and processed further on this smaller block size level. The difference with H.264/AVC is that these block sizes became of a wider size range. With H.264/AVC, block sizes varied between 4x4 and 16x16. HEVC provides the flexibility to divide the image in block sizes from 4x4 up to 64x64, but in essence, still block division is applied. A hybrid video compression standard indicates that these blocks can be predicted by means of motion compensation and that the residue after prediction is transformed to a frequency based domain.

On a higher abstraction level, these blocks can be grouped together in slices which can be independently decoded by an HEVC decoder. The downside of slices is that compression efficiency is reduced because slice independability reduces the information to predict from. After decoding the slices, these are put together forming a video picture. Similar to H.264/AVC, slices provide the possibility to decode parts of the picture in parallel and reduce error in the picture when network packets are lost at a cost of compression efficiency. Although slices provide robustness against packet loss, in this paper only a single slice per picture will be considered.

As in H.264/AVC, the use of motion compensation in a picture is indicated by the picture type. With an Intra (I) predicted picture no motion compensation is applied. Additionally, unidirectionally Predicted (P) and Bidirectionally predicted (B) pictures enabling motion compensated prediction are again part of the standard.

Because effects of packet loss are investigated in this paper, the properties of I-, P-, and B-pictures on packet loss should be considered. Because coding efficiency within Pand B-pictures is mainly obtained by temporal prediction, distortions caused by packet loss are propagated with these prediction types. Intra predicted pictures have the ability to stop the propagation of these errors, but only when applied as random access points. With the development of HEVC, compression improvements are obtained both for the intra (I) predicted and the inter (P, B) predicted pictures [11]. When only considering intra prediction in a video stream, 25% Bjøntegaard Delta (BD) [12] bitrate decrease can be obtained with HEVC compared to H.264/AVC. Only encoding the first picture in the video sequences as an I-picture followed by inter predicted pictures results in -44% BD-rate decrease for HEVC compared to H.264/AVC. From these numbers, it can be concluded that the performance of inter prediction improved more compared to the performance of intra prediction during the HEVC development process. This compression improvement difference is important for observations made from subjective evaluations, but to get in more detail about these results, first the test set up will be explained next.

3. METHODOLOGY

To evaluate the packet loss effects on HEVC encoded video streams, a test methodology as described in [10] is followed.

A set of eight sequences is selected (see Table 1) covering a large variety in the Spatial perceptual Information (SI) and Temporal perceptual Information (TI) space [13]. This to ensure that the set of sequences is representative for video transmitted over an IPTV network. In this paper, the focus is on High Definition content and therefore all sequences are of a 1920x1080 pixels resolution. All sequences are 25 frames per second and have an exact length of 10 seconds.

In the next step of the process, these sequences are encoded with both an H.264/AVC encoder (Joint reference Model JM 16.1 [14]) and an HEVC encoder (HEVC reference Model HM v4.0 [15]). For both compression standards, the same compression configurations are applied to make a comparison possible. To enable random access and recovery from

 Table 1. Test sequences

Sequence	Description
basketball	Camera pans and zooms to follow the action.
BBB	Big Buck Bunny. Computer-Generated Imagery.
cheetah	Camera pans to follow the cheetah.
ED	Elephants Dream. Computer-Generated Imagery.
foxbird3e	Cartoon. Fast camera pan with zoom.
purple4e	Many small objects moving in a circular pattern.
rush hour	High depth of focus. Fixed camera.
SSTB	Sita Sings the Blues. Slight camera zoom in.

 Table 2. MOS and standard deviation with scenarios showing significant correspondence in bold.

Loss identification		H.264/AVC MOS		HEVC MOS		
			Avg	Std Dev	Avg	Std Dev
IPPPP	Ι	BEGIN	3.98	0.71	3.71	0.75
IPPPP	2P	BEGIN	2.88	0.79	2.23	0.76
IPPPP	2P	MIDDLE	3.36	0.75	2.44	0.79
IPPPP	Р	END	4.53	0.49	4.32	0.61
IPPPP	2P	END	4.15	0.75	3.93	0.77
IPBPB	Ι	BEGIN	3.41	0.73	3.06	0.95
IPBPB	Р	END	3.71	0.79	3.53	0.79
IPBPB	В	END	4.89	0.31	4.86	0.36
IPBBB	Ι	BEGIN	3.99	0.74	2.49	0.91
IPBBB	2B	END	4.61	0.48	4.80	0.45
	no los	SS	4.97	0.24	4.96	0.21

packet loss error drift, Intra pictures are included every 15th or 16th picture depending on the number of B-pictures in the video stream. Configurations varying the number of B-pictures come in three tested varieties. In a first configuration only P-pictures are included between the interval of I-pictures (IPPPPPPPPPPPP). This repeating number of pictures starting with an I-picture is called a Group Of Pictures (GOP) from now on. A second configuration is made with a B-picture after every P-picture (IPBPBPBPBPBPB). As a third variation, three B-pictures are inserted after a P-picture (IPBBBPBBBPBBBB). For all configurations, every picture is encoded as a single slice. So, by combining the eight source video sequences with these three configurations for every sequence, 24 test sequences are created.

After video stream generation, impairments are introduced on the video stream by means of the open-source media streamer called Sirannon [16]. All impairments are on a slice level and no impairments are introduced in the first and last two seconds. To cover a large variety of slice drop scenarios with a limited number of sequences to evaluate, a limited set of 13 loss scenarios are created. From this set, 10 loss scenarios are chosen with a fixed configuration (see Table 2) and three scenarios contain more flexibility (see Table 3). In the first column of these tables, the encoding configuration on which the slice loss scenario is applied is abbreviated. The next column indicated the picture type of the lost picture. It is also indicated when two instead of one picture is lost in the video stream. From the three scenarios in Table 3, randomly one or two P- or B-pictures are removed. This to create more variety

 Table 3. Three tested scenarios not included to test results because of P.B flexibility.

Loss identification			
IPBBB	P,B	BEGIN	
IPBBB	2P,2B	MIDDLE	
IPBBB	2P,2B	END	

Table 4. MannWhitney U test results with scenarios havingsignificant correspondence in bold.

Loss identification			Significance
IPPPP	Ι	BEGIN	0.107
IPPPP	2P	BEGIN	0.000
IPPPP	2P	MIDDLE	0.000
IPPPP	Р	END	0.008
IPPPP	2P	END	0.045
IPBPB	Ι	BEGIN	0.005
IPBPB	P	END	0.077
IPBPB	В	END	0.803
IPBBB	Ι	BEGIN	0.000
IPBBB	2B	END	0.000
no loss			0.995

in the test data and therefore create a more diverse dataset. The next column indicates where in the GOP the picture loss occurs.

So, for each compression format, H.264/AVC and HEVC, 104 sequences containing loss effects are decoded and subjectively evaluated. For decoding the impaired sequences a modified version of the compression software is used. Modifications are made to facilitate error concealment by means of the frame copy algorithm.

Subjective evaluation follows a Single Stimulus (SS) Absolute Category Rating (ACR) assessment methodology. Before starting the evaluation, color vision and visual acuity of the participants is checked by means of an Ishihara test and a Snellen chart respectively. Then, three training sequences are presented to make the participant familiar with possible errors introduced in the video sequences. To restrict the time for subjective evaluation to 20 minutes, the impaired sequences are split into four datasets. All evaluations are done on a 40 inch full HD LCD screen with participants sitting at four times the display height from the screen. For every dataset exactly 24 valid participants evaluated the quality by indicating the quality on a scale from zero to five. To verify the validity of the participant's results, a post-experiment screening as described in annex V of the Video Quality Experts Group (VQEG) HDTV report is performed.

4. RESULTS

To evaluate the consequences of HEVC compression instead of H.264/AVC compression on the subjective quality after picture loss, first the MOS from the subjective evaluation are analysed. Afterwards, results from applying an H.264/AVC impairment detection model on the HEVC compressed sequences are discussed.

4.1. HEVC quality assessment

Average and standard deviation results from the subjective comparison between H.264/AVC and HEVC with picture loss is summarized in Table 2. In this table, only results from loss scenarios that were fixed are visualised. From the loss scenarios in Table 3, where a random choice between P-picture or B-picture losses could be made, no average and standard deviation results are listed. This because a direct comparison does not give relevant results. In Table 2, three scenarios are marked in bold, because from these three, significant correspondence is verified by means of a MannWhitney U test between both sets of results. The MannWhitney U significance of all compared scenarios are shown in Table 4. In this table, MannWhitney U significance values larger than 0.05 indicate that the distributions in both sets of subjective results are the same. In the last row of both result tables, statistics of the MOS-results of unimpaired sequences is given. From the large MannWhitney U significance, it can be derived that both test groups from the H.264/AVC and HEVC-test show the same MOS-distribution.

When further inspecting results of the impairment scenarios, it can be observed that correspondence in MOS-results is especially noticeable when the MOS is high. On lower MOS, the difference between HEVC MOS-results and H.264/AVC MOS-results increases significantly. Certainly when B-pictures are incorporated, differences between subjective results increase. For example, when looking at the IPBBB scenario where an I-picture is lost, a difference of 1.5 MOS-difference (3.99-2.49) can be observed. From this observation it can be concluded that because of the more efficient temporal prediction of HEVC more visual artefacts can be perceived caused by the introduced frame loss. When a picture loss is introduced resulting in a significant error propagation in time, the improved temporal prediction of HEVC makes the error more visible than when H.264/AVC is used.

4.2. HEVC model

In [10], a subjective model was created indicating network impairment visibility for H.264/AVC encoded video streams. The proposed bitstream-based model extracts different parameters from the video stream to base a visibility decision on. The parameters which are important for a high level model are enumerated in Table 5. All these parameters can be extracted from the received video stream without the need for deep packet inspection. With these bitstream parameters and with the results from subjective evaluation of the sequences, a decision tree is generated using the Waikato Environment for Knowledge Analysis (WEKA). The purpose of the resulting decision tree is to classify the visibility of packet loss. In this research, packet loss is marked as visible when 75% of the

 Table 5. Video stream parameters.

Parameter	Description
contentclass	Sequence content classification.
imp_pic_type	Picture type (I, P, or B) of lost picture.
imp_in_gop_pos	Temporal location of lost picture.
only_b_slice_drops	Only B-picture drops.



Fig. 1. High-level parameter-based model developed for H.264/AVC impairment visibility detection [10].

participants rated the overall quality of the video as visually degraded.

Based on the subjective evaluation of the H.264/AVC video streams, two decision trees were created. A first objective model is solely based on high-level bitstream parameters and is schematically represented in Fig. 1. The depicted decision tree differs from the decision tree in [10] because parameters considering multiple slices are removed from the decision tree. On H.264/AVC encoded video streams, this model resulted in a 83.1% prediction accuracy. This prediction accuracy indicates for how many test scenarios the model predicts visibility of packet loss corresponding to the evaluations of the test group. When applying the same model on HEVC encoded content, a smaller but still significant visibility prediction with 81.7% accuracy can be obtained. This result indicates that high level models designed for H.264/AVC video compression can show significant results on HEVC compressed videos.

With the construction of the second model from [10], a content classifier is taken into account. The result of the classifier which is based on the SI and TI information of the test

Table 6. Content classification based on the amount of motion and spatial detail into four classes.

Content Class	Characteristics	Sequences
A	low motion,	BBB, rush hour
	low spatial detail	
В	high motion,	cheetah, foxbird3e
	medium spatial detail	
С	high motion,	basketball, purple4e
	high spatial detail	
D	low motion,	ED, SSTB
	high spatial detail	



Fig. 2. High-level parameter-based model incorporating content classification developed for H.264/AVC impairment visibility detection [10].

sequences is illustrated in Table 6. With this extra information included, a model is created as illustrated in Fig. 2. With this model, improved impairment detection of H.264/AVC content could be obtained. Results indicate visibility prediction accuracy of 86.1%. Applying this model on HEVC content reduces the prediction accuracy to 76.0%. Contrary to the previous results, prediction accuracy decreases instead of increases when applying the content classification based model on HEVC. Therefore, concluding that every H.264/AVC model is directly valid for HEVC is certainly incorrect.

When considering the parameters as described in the second model, a new model was created based on the subjective results from the HEVC evaluation. The resulting model is indicated in Fig. 3. The accuracy of HEVC impairment visibility detection is increased with this model from 83.1% to 83.7%. When compared to the H.264/AVC model, the same decisions can be taken for losses of I-pictures and B-pictures. For P-pictures, a small simplification of the model can be observed. Similar to H.264/AVC, a loss of a P-picture at the beginning of the GOP is still indicated as visible. For P-picture losses in the middle of the GOP, the increased visibility of picture loss in HEVC can be observed. Irrespective of the content, all picture losses on this position are marked as visible. On the contrary, when a P-picture loss occurs at the end of the GOP, reduced visibility can be observed in the created HEVC model.

Although certain impairment detection models created for H.264/AVC can be applied on HEVC, it can not be assumed that every model can be ported to HEVC compression without slight modification.

5. CONCLUSION

In this paper, we show that impairment visibility slightly increases when changing from H.264/AVC to HEVC video compression. This is caused by the increased performance of temporal prediction between video pictures. More data from previous pictures is used for the prediction of following pictures propagating visual artefacts more through the compressed video



Fig. 3. High-level parameter-based model incorporating content classification modified for HEVC impairment visibility detection.

stream. As a consequence, this results in increased visibility of packet loss introduced on the video stream. Although impairment visibility increases, a high level impairment detection model designed for H.264/AVC performs similarly when applied on HEVC compressed video content. Although similar performance was obtained for this high level model, when improving the H.264/AVC model by incorporating content classification, performance degradation of the model was observed when applied on HEVC content. Slight modification of the model considering content classification could increase the performance again. Therefore, it is advised to be cautious when applying an H.264/AVC impairment detection model on HEVC.

6. ACKNOWLEDGEMENT

The research activities described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research–Flanders (FWO–Flanders), and the European Union.

7. REFERENCES

- G.W. Cermak, "Consumer Opinions About Frequency of Artifacts in Digital Video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 336–343, April 2009.
- [2] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Transactions on Broadcasting*, vol. PP, no. 99, pp. 1, 2011, [to be published].
- [3] ITU-T Recommendation J.247, "Objective perceptual multimedia video quality measurement in the presence of a full reference," International Telecommunication Union (ITU), 2008.
- [4] ITU-T Recommendation J.246, "Perceptual visual quality measurement techniques for multimedia services

over digital cable television networks in the presence of a reduced bandwidth reference," International Telecommunication Union (ITU), 2008.

- [5] F. Yang, S. Wan, Q. Xie, and H. R Wu, "No-reference quality assessment for networked video via primary analysis of bit stream," *IEEE Transactions on Circuits* and Systems for Video Technology, vol. 20, no. 11, pp. 1544–1554, November 2010.
- [6] M.N. Garcia and A. Raake, "Parametric packet-layer video quality model for IPTV," in 10th International Conference on Information Sciences Signal Processing and their Applications (ISSPA), May 2010, pp. 349–352.
- [7] N. Staelens, I. Sedano, M. Barkowsky, L. Janowski, K. Brunnström, and P. Le Callet, "Standardized toolchain and model development for video quality assessment - the mission of the joint effort group in VQEG," in *Third International Workshop on Quality* of Multimedia Experience (QoMEX), September 2011.
- [8] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems* for Video Technology, vol. 13, no. 7, pp. 560–576, Jul. 2003.
- [9] B. Bross, W.-J. Han, J.-R. Ohm, G.J. Sullivan, and T. Wiegand, "WD4: Working Draft 4 of High-Efficiency Video Coding," *JCTVC-F803*, Torino, Italy, Jul. 2011.
- [10] N. Staelens, G. Van Wallendael, K. Crombecq, N. Vercammen, J. De Cock, B. Vermeulen, R. Van de Walle, T. Dhaene, and P. Demeester, "No-Reference Bitstream-based Visual Quality Impairment Detection for High Definition H.264/AVC Encoded Video Sequences," *IEEE Transactions on Broadcasting*, 2012, Accepted for publication.
- [11] B. Li, G. J. Sullivan, and J. Xu, "Comparison of Compression Performance of HEVC Working Draft 4 with AVC High Profile," *JCTVC-G399*, Geneva, Switzerland, Nov. 2011.
- [12] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," *document VCEG-M33 of ITU-T Video Coding Experts Group (VCEG)*, Apr. 2001.
- [13] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union (ITU), 1999.
- [14] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "Doc. JVT-AE010: H.264/14496-10 AVC Reference Software Manual," Tech. Rep., MPEG / ITU-T, Jun. 2009.

- [15] K. McCann, B. Bross, S.-i. Sekiguchi, W.-J. Han, "HM4: High Efficiency Video Coding (HEVC) Test Model 4 Encoder Description," *JCTVC-F802*, Torino, Italy, Jul. 2011.
- [16] A. Rombaut, N. Staelens, N. Vercammen, B. Vermeulen, and P. Demeester, "xStreamer: Modular Multimedia Streaming," in *Proceedings of the seventeenth ACM international conference on Multimedia*, 2009, pp. 929–930.