

Subjective Assessment for Standard Television Sequences and Videotoms – H.264/AVC Video Coding Standard

Łukasz Trzcianowski

Faculty of Electronics and Information Technology, Warsaw University of Technology, Warsaw, Poland

Abstract—This paper presents comparison of videotoms and standard television sequences in terms of image distortions and perceived subjective quality affected by H.264/AVC compression with changed bit rate. Results from initial tests, performed as laboratory exercise can be a reference to show scale of diversity in both level of degradation and Mean Opinion Score (MOS) evaluation. Results and comments included in this paper give overview on the codec influence on videotoms and can suggest approach for further tests and experiments.

Keywords—*digital television, IPTV, Quality of Experience, Quality of Service, videotoms.*

1. Introduction

The rapid growth of multimedia applications in the recent years has raised needs to provide services to clients in efficient way. In this case, not only network performance should be taken into account but also coding methods, especially those considered as standards. They are still under development in terms of improvements for compression efficiency while preserving same quality. Good example that comes to mind is H.265/HEVC technique created to fully support Ultra High Definition Television (UHDTV) and video resolutions up to 8192×4320 [1]. In fact, this standard is too new to be widely used in commercial systems for providing television services especially in Standard Definition Television (SDTV) and for such a video format it is reasonable to use H.264/AVC method [2], [3] in all the quality assessment tests.

Obviously coding process is only one factor that can affect video quality perceived by viewers. Important is how video signal is prepared, transmitted and received and all those parts should be always covered in video quality approximation. In telecommunication services provided over the IP network, usually parametric models are used in the planning process in order to meet Quality of Service (QoS) requirements, but for video applications it is difficult to create general recommendation that can be applied to all possible implementations and conditions. The most important issue is to find out what is the real relationship between parameters corresponded to coding, transmission, and receiving methods and user satisfaction from provided service.

Current works on this topics keep focus on solutions associated with specific conditions considered in experiment process [4]–[7]. Basing on this and even considering existing ITU-T standards for Quality of Experience (QoE) [8] or P.NAMS, P.NBAMS [9], [10] that refer to packet and stream layer, it is difficult to talk about parametric models for IPTV services.

The author's main idea is to use videotoms as simple video sequences in order to create general approach to parametric model creation that could be moved to real TV materials. During tests focused mainly on network conditions [11], [12] and in further experiments it was found that H.264/AVC coding impact is not visible for simple videotoms sequences, whereas in standard television sequences the influence is significant. To verify that and to check what is the scale of degradation between two considered types of video the author decided to do firstly initial subjective tests for coding bit rate without including network parameters usage. There is a lot of papers that cover H.264/AVC coding methods [13]–[15] but not for simple sequences like videotoms.

2. Video Sequences

Video sequences used for this paper purposes can be divided into two groups: standard television video sequences and videotoms.

Difference between both groups is significant. The first group includes various video probes usually extracted from the TV programs source materials, whereas the second one contains sequences created based on human visual perception characteristic. Videotom's definition was introduced in another author's publication [11] and it refers to simple, well-defined and known to users video materials. In order to create such sequences both all image elements and dependencies between them should be taken into consideration. In this case, video pictures should keep constant form and organization. Each irrelevant or abstract information should be removed due to limitations in perceiving process. It is important to take into account also such elements like contrast, brightness, details level, dynamism, and diversity. Videotoms do not fit to television materials

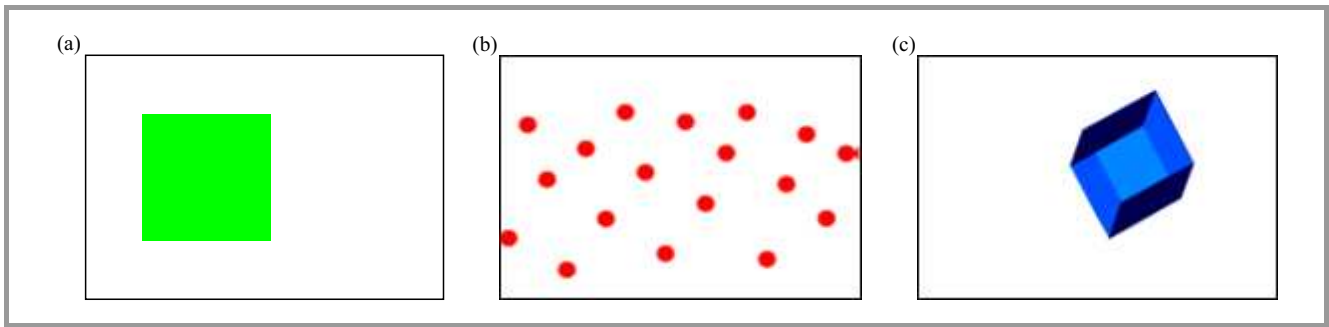


Fig. 1. Videotoms: (a) Green rectangle, (b) Red balls, (c) Blue cube.



Fig. 2. Television sequences: (a) Suzie, (b) Mr. Fins, (c) Cheerleaders.

Table 1
Videotoms description

Name	Duration [s]	FPS [fps]	Resolution [pixels]	Description
Green rectangle	10	25	720 × 576	Low motion sequence with small number of details (videotom)
Red balls				Average/high motion sequence with average number of details (videotom)
Blue cube				High motion sequence with small number of details (videotom)

Table 2
Television sequences description

Name	Duration [s]	FPS [fps]	Resolution [pixels]	Description
Suzie	10	25	720 × 576	Low motion sequence with small number of details
Mr. Fins				Average/high motion sequence with average number of details
Cheerleaders				High motion, detailed sequence with background

because of simplification due to very strong adaptation to human seeing process described in many studies [16], [17]. The benefit from their usage is related to easy analysis of the subjective assessment process because of the fact that distortions are easier to notice for users than in the standard television sequences. In studies and tests regarding their diversity, not only transmission should be considered, but also scale and behavior in terms of standard coding methods.

3. Test Conditions

3.1. Test Materials – Reference and Processed Probes

To verify scale of diversity in the degradations caused by H.264/AVC coding between standard television materials and videotoms six sequences were selected. First three were downloaded from “Consumer digital video library” [18] and other were created using Macromedia Flash Professional application in order to be adapted to the nature of human visual perception and according to Young-Helmholtz theory of trichromatic color vision in terms of receptors engagement. All the sequences are presented in Figs. 1 and 2, together with their parameters shown in the Tables 1 and 2.

Presented reference probes were used in the H.264/AVC coding process to create test materials. For this process author decided to use open implementation x.264 installed in the VirtualDub application. Settings setup was in line with ITU-T H.264 standard [2] and with Video Quality Experts Group (VQEG) recommendations – main profile, level 3.0. Coder bit rate was changed in the range of the following values: 3000, 2000, 1000, 500, 250, 100, and 64 kb/s using QP parameter. The subject of this paper focuses on H.264/AVC coding method and influence of network parameters were intentionally omitted.

3.2. Test Method

As a test method, Double Stimulus Impairment Scale (DSIS) was selected – approach and conditions are described in details in ITU-R recommendation BT.500-11 [19]. In this screening process two sequences are shown to assessor in pairs: first one is the reference, second one is impaired (after processing). Important is that viewer is informed about the order and after their playback, he is asked for the quality evaluation using impairment scale: 5 – imperceptible, 4 – perceptible, but not annoying, 3 – slightly annoying, 2 – annoying, 1 – very annoying. To proceed with tests MSU Perceptual Video Quality tool was used [20], where each single projection contains 3 parts: reference sequence, 3s grey area, and impaired sequence.

3.3. Testers

Tests were executed by students (15 persons) in similar age group during laboratory exercise. They were trained on test procedures, used methods, metrics, and tools.

The objective for them was not only to assess the quality but also to provide information about observed distortions. In many cases, it is recommended to engage experts in subjective tests but it is expensive what creates difficulties especially in the initial tests. In this paper author assumes that trained non-expert testers can produce similar results if instructions and guidance is provided in a proper way (short training session with examples).

3.4. Results – MOS, Confidence Intervals

After tests execution MOS was computed for each particular test condition as an average of obtained results as:

$$MOS_k = \frac{1}{N} \cdot \sum_{i=1}^N MOS_{ik}, \quad (1)$$

where: N – number of testers after outlier removal, MOS_{ik} – score assigned by tester i to test condition k , MOS_k score for test condition k .

To measure the estimate reliability based on a sample of population (15 persons) confidence intervals of estimated mean were calculated. Results show the relation between estimated mean values and entire mean values of the entire population. Due to small number of students and assuming 95% confidence level, intervals for the mean subjective scores were computed by using Student-T distribution as:

$$\delta = t_{(1-\frac{\alpha}{2})} \cdot \frac{S}{\sqrt{N-1}}, \quad (2)$$

where: δ – confidence, $t_{(1-\frac{\alpha}{2})}$ – t value associated with given significance level α for a two-tailed test, $N-1$ – degrees of freedom, where N is the number of observations in the sample, S – estimate standard deviation of the sample of observations.

4. Results and Conclusions

In this section test results are considered as a comparison two sets of mean MOS scores - one for standard sequences and the second one for videotoms. Figures 3 and 4 show, for each video content, mean MOS values and confidence intervals across changes in the H.264/AVC codec bitrate. Received subjective rates present what is the degradation level and impact on the quality for both types of used materials. MOS results for videotoms are comprised between 2.75 and 5, but for television sequences they span over quality levels entire range. Obviously extremely low coding bit rate negatively affect video quality in all tested sequences, but in case of videotoms even by the smallest bitrate values video pictures are still readable. It is worth to comment on confidence intervals that inform about results reliability. They are definitely more wide for television sequences and it is not possible to identify for which bitrate values the less precise estimates (greater level of variance) can be expected because this depends on particular video sequence. For videotoms, confidence intervals are relatively small, wider

only by extreme values of coding bitrate. It is also significant to indicate that for television sequences MOS values decrease faster, starting by 2000 kb/s. This behavior is in direct relation to TV materials where usually we can expect more details and dynamism and background layers in the video pictures. Standard H.264 mechanisms like motion compensation, entropy coding and inter/intra picture prediction are easier for videotoms and maybe it is reasonable to Basic Line profile instead of Main profile, but the objective for these tests is to have same conditions for both types of sequences.

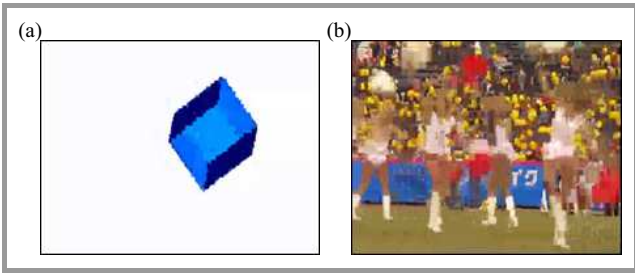


Fig. 5. Blue cube (videotom), 64 kb/s (a), Cheerleaders (TV), 250 kb/s (b).

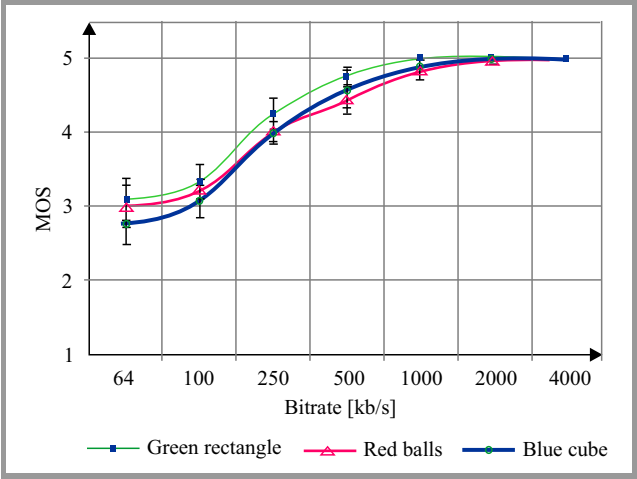


Fig. 3. Mean MOS score for videotoms.

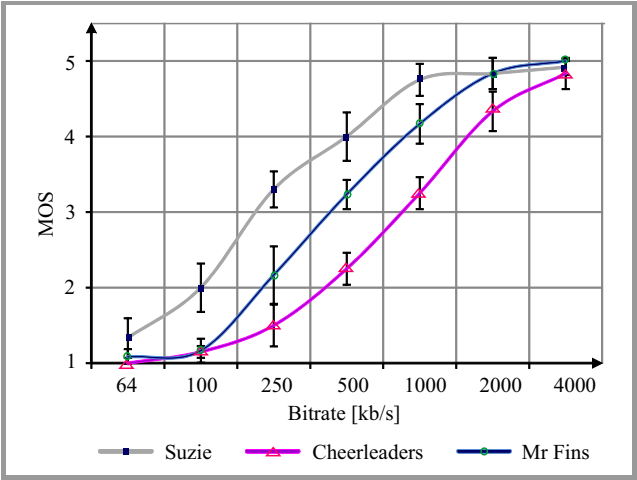


Fig. 4. Mean MOS score for standard sequences.

Regarding to the scale of diversity, by coding bitrate 250 kb/s the MOS value for “Suzie”, “Mr. Fins” and “Cheerleaders” is included between 1.5 and 3.5, whereas for videotoms it is still acceptable (above 4). Figure 5 illustrates what is the scale of difference in degradation level. Observed distortions show that in many cases impact of coding bitrate seems to be negligible for videotoms in comparison to effects in TV sequences. Apparently an error concealment mechanisms works for such simple sequences causing small number of visible degradations and in effect

good quality assessed by viewers. Provided in this article results and comments should be treated as initial experiment, which should be extended in order to draw final conclusions and present mathematic description for the relations. However, results are helpful to indicate difference scale between considered types of video sequences.

5. Summary

This paper is a part of work executed in Institute of Telecommunication at Warsaw University of Technology in terms of multimedia services and quality assessment. Results present how decreased coding bitrate affects video quality perceived by viewers as well as the scale of difference in the level of degradations between videotoms and standard television sequences. The most frequently occurred degradations are same for both types of video material and they are mainly related to blocking and flicker effects and to changes in shapes and colors. Because of the complexity and overall characteristic of TV sequences are more sensitive to negative impact of low bitrate values. For videotoms H.264/AVC coding mechanism seem to work very efficiently and degradations are visible only by lowest bitrate values. Presented in the scope of this work results have to be extended with additional experiments considering influence of network changes according to earlier author’s work [11], [12]. To do that firstly it is required to produce more videotoms sequences grouped into several sets based on their characteristic to compare them in the next step to television sequences with the same settings. It is also needed to engage more viewers in subjective tests in order to allow on more detailed statistic analysis for the results. For objective tests both simple metrics and more complex like Perceptual Video Quality (PVQ), eMOS [21] can be used as point of reference. Final results can allow author to determine functional relationship between considered types of sequences to create proper approach to parametric model creation.

References

[1] L. Trzcianowski, “Videotoms in objective and subjective quality tests of video”, *J. Telecommun. Inform. Technol.*, no. 1, pp. 27–31, 2014.

- [2] Ł. Trzcianowski and S. Kula, "Approach to QoS/QoE monitoring and parametric model for IPTV service", *Scient. J. Telecommun. Electron.*, vol. 16, no. 261, pp. 41–59, 2012.
- [3] J. Asghar, "Preserving video quality in IPTV networks", *IEEE Broadcast Technol. Soc.*, vol. 55, 2009.
- [4] A. Raak *et al.*, "T-V-model: parameter-based prediction of ITV quality", in *Proc. IEEE Int. Conf. Acoust. Speech and Sig. Process. ICASSP 2008*, Las Vegas, NV, USA, 2008, pp. 1149–1152.
- [5] K. Yamagishi and T. Hayashi, "Parametric packet-layer model for monitoring video quality of IPTV services", in *Proc. IEEE Int. Conf. Commun. ICC 2008*, Beijing, China, 2008, pp. 110–114.
- [6] M. Vranjes, S. Rimac-Drlje, and D. Zagar, "Subjective and objective quality evaluation of the H.264/AVC coded video", in *Proc. 15th Int. Conf. Syst. Sig. and Image Process. IWSSIP 2008*, Bratislava, Slovak Republic, 2008, pp. 287–290.
- [7] Z. Milicevic and Z. Bojkovic, "Subjective video quality assessment in H.264/AVC video coding standard", in *Proc. 19th Telecommun. Forum TELEFOR 2011*, Belgrade, Serbia, 2011, pp. 1183–1186.
- [8] F. De Simone *et al.*, "Subjective quality assessment of H.264/AVC video streaming with packet losses", *EURASIP J. Image Video Process.*, vol. 1, 2011.
- [9] S. S. Hemami and A. R. Reibman, "No-reference image and video quality estimation: applications and human-motivated design", *Sig. Process.*, vol. 25, no. 7, pp. 469–481, 2010.
- [10] C. Ware, *Information Visualization: Perception for Design*. Morgan Kaufmann, 2013.
- [11] L. Zhaoping, *Understanding Vision: Theory, Models, and Data*. OXFORD University Press, 2014.
- [12] "Advanced video coding for generic audiovisual services", ITU-T Rec. H.264, 05/2003.
- [13] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard", *IEEE Trans. Circ. Syst. for Video Technol.*, vol. 13, no. 7, pp. 560–576, 2003.
- [14] "High efficiency video coding", ITU-T Rec. H.265, 04/2013.
- [15] "Methodology for the subjective assessment of the quality of television pictures", ITU-R Rec. BT.500-11, 06/2002.
- [16] "Parametric non-intrusive assessment of audiovisual media streaming quality", ITU-T Rec. P.1201, 10/2012.
- [17] "Parametric non-intrusive bitstream assessment of video media streaming quality", ITU-T Rec. P.1202, 10/2012.
- [18] "Quality of experience requirements for IPTV services", ITU-T Rec. G.1080, 12/2008.
- [19] "Consumer digital Video Library"– CDVL [Online]. Available: <http://www.cdvl.org>
- [20] MSU Perceptual Video Quality Tool [Online]. Available: http://www.compression.ru/video/quality_measure/subjective_metrics_info_en.html
- [21] "Perceptual Video Quality (PVQ)" [Online]. Available: <http://www.tek.com/products/video-test/sentry-solutions/perceptual-video-quality.html>



Łukasz Trzcianowski graduated from the Faculty of Electronics and Information Technology at Warsaw University of Technology (M.Sc., 2012). At present, he continues his work as Ph.D. candidate (Institute of Telecommunications, Warsaw University of Technology). The main focus in his research is given on telecommu-

nication services and parametric models.

Email: ltrzczianowski@tele.pw.edu.pl

Institute of Telecommunication

Warsaw University of Technology

Nowowiejska st. 15/19

00-665 Warsaw, Poland