Towards using entertaining 3D video sequences in entertainment-oriented subjective experiments

Dawid Juszka and Zdzisław Papir
Department of Telecommunications, AGH University of Science and Technology
Kraków, Poland
email: juszka@kt.agh.edu.pl

***Abstract*—Both the wide access to Video-on-Demand services and the large availability of tag information enable today's users to make conscious decisions on the video content they watch. This signifies a dramatic change in user behaviour compared to previous generations, whose choice was constrained by TV listings, which further restricted the time of viewing. Most subjective video quality studies use short-duration sequences that are specifically designed to use encoding algorithms and do not consider the subjective characteristics of the content as variables. For these reasons, the user's motivations to watch video content should be included in Quality of Experience measurement methods. This paper shows that subjective assessment of personal interest in video content, its visual attractiveness and quality of 3D effects do not correlate with objective parameters usually used for test sequence selection: spatial perceptual information, temporal perceptual information and mean scene-cut density.**

I. Introduction

Observations of trends in popular VoD services prove that if the user is highly interested in the content of a video he or she accepts that it may not be the best visual quality, but watches it regardless. So, the decision to watch the video is based not only on an acceptable level of visual quality, but also on a subjectively assessed level of content desirability [1][2].

The selection of video sequences for subjective experiments is crucial for the reliability of test results. Currently recommended techniques for choosing such sequences are collected in [3]. When designing a subjective experiment, researchers take two actions to minimize any impact of artistic, aesthetic and storyline aspects on ratings: they instruct subjects not to base their opinion on the content of the scene or the quality of the acting and therefore select the most neutral video content as a source material. Unfortunately, results inevitably include both the clip's artistic and technical qualities. It is quite surprising that entertainment-oriented subjective experiments are conducted with source sequences intentionally selected to be devoid of basic entertaining features. We believe that there is an urgent need to begin using entertaining test sequences instead of neutral content (which is usually boring) to get results relevant to the nature of services they relate to.

One main and inherent feature of entertainment video content is the occurrence of scene-cuts which are a natural consequence of making footage attractive by using video editing. Neglecting the existence of scene-cuts in such type of video was the prevalent stance expressed in VQEG throughout the 1990s. The current strategy is to select footage with scene-cuts [4]. In this work we select only sequences with abrupt scene transitions.

In this paper we present our research in terms of video sequence selection for entertainment-oriented QoE subjective testing for 3D video needs. We have selected and subjectively measured three characteristics of 3D video content: interestingness, visual attractiveness and 3D effect experience. We then studied correlations between them and objective parameters recommended by ITU to select source sequences for subjective experiments i.e. detailedness of video image, its dynamism and additionally our own parameter: scene-cut density.

II. Experiment design and setup

The aim of the experiment is to measure the subjective experience of 3D video sequences in regards to three aspects: interestingness (INT), visual attractiveness (ATR) and 3D effect impression (3DE), in addition to commonly used objective parameters SI and TI.

1) Source content: The source video material for our experiment was cut from 9 movie productions (7 feature films and 2 documentaries) recorded on Bluray-3D discs (*The Amazing Spider-Man*, *Born to be wild*, *Drive Angry*, *Hugo*, *Man of Steel*, *Pirates of the Caribbean: On Stranger Tides*, *Polar Bears: Ice Bear*, *The Great Gatsby*, *The Hobbit: An Unexpected Journey*).

The movies were extracted from Bluray-3D discs with the highest possible quality (H.264 AVC (part 10), avc1, 24 fps, 4:2:0 yuv). We decided to select a scene only if we could recognize a logical course of action within the bounds of $30 \pm 2s$. Consequently, we selected 96 scenes from the acquired material and 5 sequences for the purpose of training. Next, all 96 sequences were ordered randomly and divided into 3 equinumerous sets.

Objective parameters were calculated for each sequence:

- The spatial perceptual information (SI) and the temporal perceptual information (TI) [5].
- Mean SI value for whole sequence:

$$SI_{mean} = \frac{\sum std_{space}[Sobel(F_n(i,j))]}{N}. \quad (1)$$

- Mean TI value for whole sequence:

$$TI_{mean} = \frac{\sum std_{space}[F_n(i,j) - F_{n-1}(i,j)]}{N}. \quad (2)$$

- Scene-cut density:

$$d = \frac{\sum [r_m - r_{m-1}]}{M}. \quad (3)$$

where: F_n – the luminance plane of video frame at time n; $std_{space}(\bullet)$ – standard deviation over the pixels; $Sobel(\bullet)$ –

a Sobel filter; r_m – the frame number where m-th scene-cut occurs; N – total number of frames in a sequence; M – total number of scene-cuts in a sequence.

Calculations were made only for a left view and values for scene-cut frames were excluded from the measurement. Also, for this reason, two additional parameters SI_{mean} and TI_{mean} were calculated to depict the characteristics of source sequences. We found objective selection criterion based on *criticality* [3] incongruent, because despite the wide variety of SI and TI scores – none of sequences exceeded the 1.87 *criticality* score.

2) Test environment: The test environment was prepared according to recommendations: ITU-R BT.2021, ITU-R BT.500 and ITU-T P.910. As a display, a 42" 3D plasma screen PANASONIC TX-P42GT30 was used. During the experiment subjects were equipped with shutter glasses.

3) Subjects: Each of the 28 subjects (14 men and 14 women) was a naive observer, selected by a recruitment agency which pre-checked their stereo vision acuity and colour vision, using the RANDOT Stereo Test and the Ishihara colour vision test, respectively. Positive results were a necessary condition for taking part in the experiment.

4) Methodology description: Each subject took part in the experiment alone. Due to the fact that in this experiment we did not ask subjects to assess the quality of stimuli, but rather their opinions on content, the Absolute Category Rating (ACR) scenario was chosen. After each sequence, the subject had 15 seconds to answer three questions on the Likert scale (1-5): 1) How interesting was the presented movie clip?; 2) How visually attractive was the presented movie clip?; 3) How intensely did you experience the 3D effect in this movie clip? In the instructions for the experimental task, the subjects were given an interpretation of each question. Firstly, the subject was instructed to assess the interestingness with regards to the storyline and the type of presented action. Secondly, the subject was asked to assess the visual attractiveness in an artistic and aesthetic sense. And thirdly, the subject was instructed to assess the effect of the 3D on a scale starting from "I didn't notice the 3D effect" (as 1) to "the 3D effect was incredible" (as 5). At the end of the test subjects were asked to fill out a short questionnaire to gather information about their overall attitude towards the task and to fill in a form of the NASA Task Load Index (NASA-TLX).

III. Data analysis and results

Firstly, in the post-screening stage one subject was identified as an outlier and was excluded from further analysis.

Correlation coefficients were calculated between subjective scores (i.e. INT, ATR and 3DE) and objective parameters (i.e. SI, SI_{mean}, TI, TI_{mean}, d). Results are statistically insignificant ($p > 0,05$) only for three pairs. Correlation coefficients for other pairs are low but statistically significant (values in bold) and are presented in Table I. This suggests that objective parameters usually calculated for source video sequence selection needs (SI and TI) cannot be used alone to estimate if video content is visually attractive or interesting for subjects. Similarly, parameters such as SI_{mean}, TI_{mean} cannot be used for this purpose.

An unmissable observation is the high correlation between visual attractive scores and 3D effect experience scores. The reason for such a situation may be the difficulty in assessing these two experiences separately. Another interesting observation can be made on the basis of the NASA-TLX questionnaire results. Low scores (mean score approx. 2.7) confirm that the experimental task was not demanding for participants and justifies extending the total number of sequences in future work.

TABLE I: Correlation coefficients between objective parameters and subjective evaluation

	INT	ATR	3DE	SI	SI_{mean}	TI	TI_{mean}	d
INT	x	**0.55**	**0.36**	-0.04	-0.16	**0.28**	0.19	**-0.27**
ATR	**0.55**	x	**0.83**	**0.47**	**0.42**	**0.23**	**0.42**	**0.26**
3DE	**0.36**	**0.83**	x	**0.54**	**0.46**	**0.34**	**0.54**	**0.31**

IV. Conclusions

In the presented experiment, source sequences were drawn from 7 feature films and 2 documentaries, all in 3D. To allow the subjects a relatively short duration of testing, the length of each sequence was equal to 30 ± 2 *s*.

The aim of this study was to verify if there is a correlation between three objective characteristics: detailedness, dynamism and scene-cut density and three subjective characteristics: interestingness, visual attractiveness and 3D effect experience. Two of the objective characteristics were commonly employed parameters: SI and TI. The conducted experiment proved low correlations – not enough to recommend using such objective parameters to measure investigated subjective characteristics of source material content. Additionally, the results of this research strongly imply that the selection of sequences for subjective experiments should be driven not only by objective parameters SI and TI as recommended by ITU [5], but also by subjective content characteristics, as the latter as at least equally important as the former.

Future work will focus on measuring the influence of these subjective characteristics on quality of experience assessment scores.

Acknowledgment

This work has been supported by the Dean's Grant (agreement number 15.11.230.267) - Faculty of Computer Science, Electronics and Telecommunications, AGH UST.

References

[1] P. Kortum and M. Sullivan, "The Effect of Content Desirability on Subjective Video Quality Ratings," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 52, no. 1, pp. 105–118, may 2010.

[2] Y. Zhu, I. Heynderickx, and J. A. Redi, "Understanding the role of social context and user factors in video Quality of Experience," *Computers in Human Behavior*, vol. 49, pp. 412–426, 2015.

[3] M. H. Pinson, M. Barkowsky, and P. Le Callet, "Selecting scenes for 2D and 3D subjective video quality tests," *EURASIP Journal on Image and Video Processing*, vol. 2013, p. 50, 2013.

[4] M. H. Pinson, L. Janowski, and Z. Papir, "Video Quality Assessment: Subjective testing of entertainment scenes," *IEEE Signal Processing Magazine*, vol. 32, no. 1, pp. 101–114, jan 2015.

[5] ITU-T, "Recommendation ITU-T P.910: Subjective video quality assessment methods for multimedia applications," 2008.