3D VIDEO QUALITY OF EXPERIENCE — INFLUENCE OF SCALE AND CROSSTALK

Kjell Brunnström[1,3], Kun Wang [1,3], Börje Andrén[1], Valentin Kulyk[2], Samira Tavakoli[4], Christer Hedberg[1], Indirajith Vijai Ananth[1], Mahir Hussein[1], Mats Folkesson[2], Marcus Barkowsky[1]

[1]Acreo Swedish ICT AB, Kista, Sweden, [2]Ericsson Research, Sweden, [3]Mid Sweden University, Sweden, [4]Universidad Politécnica de Madrid, Spain, [4]IRCCyN, Nantes; France

ABSTRACT

This paper gives an overview of three recent studies by the authors on the topic of 3D video Quality of Experience (QoE). Two of studies [1,2] investigated different psychological dimension that may be needed for describing 3D video QoE and the third the visibility and annoyance of crosstalk[3].

The results shows that the video quality scale could be sufficient for evaluating S3D video experience for coding and spatial resolution reduction distortions. It was also confirmed that with a more complex mixture of degradations more than one scale should be used to capture the QoE in these cases. The study found a linear relationship between the perceived crosstalk and the amount of crosstalk.

1. INTRODUCTION

It seems that the hype of 3D TV is over, at least temporarily. The TV manufacturer tried to ride on the cinema success for 3D in hope that it would immediately carry over also in home entertainment, but this has not so far been the case. The recent hype has, therefore, calmed down and moved into slower and quieter development. 3D TV may still become big in home entertainment, but that requires that the TV manufacturer continues to bundle 3D capabilities with new TV sets. The infrastructure for delivery of content still needs to be developed, but most important, the available content needs to continue to grow steadily. Comparing with the TV transition from SDTV to HDTV, the total time required has been between 20-30 years, so the quick adoption of 3D TV was in many cases unrealistic.

Scientifically, the attention for stereoscopic 3D has advanced the field considerably. Especially in the area of Quality of Experience (QoE) for 3D video, there have been a realization that, in most cases, it is not sufficient to only study the video quality, see e.g. Lambooij et al (2011)[4], but still this is far from completely understood. This paper will give an overview of three recent studies by the authors on the topic of 3D video QoE. Two of studies [1,2] investigated different rating scales and the third the visibility and annoyance of crosstalk[3].

The paper is organized as follows. Section 2 describes the methods used in the the three studies. In Section 3 the results are presented and Section 4 summarizes the studies with the conclusions.

2. METHOD

2.1. Exp 1: Multi-scale study for 3D video coding quality

An experiment based on the NAMA3DS1 - COSPAD1 video dataset[5] was designed for comparing three different rating scales and two viewing distances. The three scales were: Visual Quality (VQ), Visual Discomfort (VD) and Sense of Presence (SP). The experimental design was based on the Absolute Category Rating (ACR) scale [6], with five levels (Excellent, Good, Fair, Poor, Bad) for the Visual Quality scale and the Sense of presence scale. The Visual Discomfort scale was based on the Degradation Category Rating scale[6] using the labels: No discomfort, Strange feeling, but not discomfortable, Slightly discomfortable, Discomfortable, and Very discomfortable.

Subjects were visually screened for visual acuity, colour vision (Ishihara) stereo acuity (Randot test) and dominating eye. The test subjects were asked to read the written instruction. To familiarize the test subjects with the procedure and the range of degradations involved in the test a training session was performed. The actual test was then done, divided into two sessions with a break in the middle. The viewing distance was changed in the break.

The test room condition complied with the ITU-R Rec. BT.500-13[7]. The lab was equipped with a Hyundai S456D 46 inch film pattern retarder stereo 3D TV. The viewers were seated at a distance of 3H (3 times the display height) from the TV, corresponding to 2.1 m in one session and at a distance of 5 H corresponding to 3.5 m in another session. The peak white luminance of TV was 177 cd/m^2 (78 cd/m^2 through passive polarized eyeglasses). The ambient illuminance level in the room was below 20 lux.

A modified version of a video player, AcrVQWin[8], developed by the authors was used to present and retrieve the responses from the test subjects.

2.1.1. Test subjects

The test subjects were of different background and age (mean 33.7, median 29, max 62 and min 18), with about 32 % females. After pre- and post-screening[9] there were 24 test subjects in total. There were mixture of Swedish and international subjects. The experiment was conducted with Swedish for native Swedish speaking subjects and English for international subjects.

2.1.2. Stimuli preparation and organization

NAMA3DS1 - COSPAD1 video dataset[2] consists of 10 different source sequences (SRC) that were processed in 11 different ways (one used as a reference was not processed), so called Hypothetical Reference Circuit (HRC)s, making a total of 110 processed video sequences (PVS) (including the references). The video duration was 16 seconds. The HRCs consisted of no processing (1 HRC), encoding-decoding (7 HRCs), resolution reduction (1 HRC), image sharpening (1 HRC) and resolution reduction plus image sharpening (1 HRC).

These sequences were prepared for the TV by subsampling the left and right view vertically and then spatial interlacing the view on every second line, so the lines from the left and right view were placed so they would overlap the corresponding polarization line on the 3D TV.

The test design was based on a 2 by 2 Latin square, where the test video stimuli were divided into two equally sized video sets A and B, as described below. The test subjects were randomly assigned into two groups, 1 and 2. Group 1 was then presented video set A first at 3H distance and followed by video set B at 5H distance, whereas Group 2 saw video set B first at 3H distance and video set A second at 5H distance.

The divided video set A and B had about the same distribution of qualities, based on an experiment performed at the Yonsei University in Korea on the same dataset using the same model of TV[9]. One SRC with all its HRCs were included as a common set. The test design was done to test different viewing distances without requiring the test subjects to conduct the test twice.

2.2. Exp 2: Multi-scale study of 3D video quality

Two subjective experiments were conducted at two laboratories in Sweden based on the same video set: one at Ericsson Research (Lab1) and one at Acreo Swedish ICT AB (Lab2). A 46” Hyundai S465D display with polarized line-interleaved pattern was used in each laboratory for displaying 3D videos in the experiment, together with polarized eyeglasses from RealD. The display was positioned 4 times of the display height (4H) in both labs.

The display was put into factory default mode except the brightness of the screen which was set to 90% i.e. 205 cd/m^2 peak level. The refresh rate was set to 50Hz. The light reduction by the passive 3D eyeglasses was measured per eye (~40%). The ambient light levels measured horizontally towards the screen and at 4H viewing distance were about 20 lux.

At Lab1, 22 naïve test subjects, participated in the experiment, where two subjects performed the test at the same time, sitting beside each other with a shielding screen in between. At Lab2, 25 naïve test subjects participated; only one subject performed the test at a time. Before the test execution, the test subjects’ vision was tested for visual acuity (Snellen Letter test), colour vision (Ishihara) and stereo acuity (Randot). One subject from each lab was rejected and thus removed from the final analysis due to inadequate results in the stereo vision test. Both labs used three voting scales for evaluating the 3D video QoE. At Lab1: “Depth Naturalness” (DN), “Video Quality” (VQ1) and “Visual Discomfort” (VD) were used with five level category scale in combination with a continuous sliding bar. At Lab2, the voting scales used in the test were “3D Realism” (3DR), “Depth Quantity” (DQ) and “Video Quality” (VQ2), with discrete five level category scale.

2.2.1. Test conditions

13 source stereoscopic video sequences (SRC), chosen from one documentary and three movies. In the scene selection scene changes were avoided. They were divided into three content types:

- *Content 1* – recorded with still camera and containing small amount of motion (standing sitting people)
- *Content 2* – recorded with still camera and containing a moderate amount of motion.
- *Content 3* – recorded using zoom and/or moving camera and containing moderate/large amount of motion.

All contents were recorded with 1920x1080 progressive (1080p) 24/1.001 fps, played in 25 fps in the test. The 1280×720 progressive (720p) videos were up-scaled to 1080p. All videos were displayed as 1920×540 per eye on the screen.

One purpose of the experiments was to verify whether 3D and 2D related properties can be assessed in the same subjective test and provide consistent results. Therefore several video processing scenarios[9], HRCs, were used to create PVSs which are listed in Table 1**.** and here classified in three groups:

- **2D -** uncompressed and compressed 2D video in full resolution and anamorphic,
- **3D -** uncompressed 3D videos with different levels of 3D quality,
- **3Denc -** compressed 3D videos at different bitrates and in Side-by-Side (SbS) format.

Table 1: List of all HRCs. The encoding bitrates considered in this test have been r01, r02, r03, r04, and r05 which correspond to 1.25, 1.5, 1.75, 2.0 and 2.5 Mbps respectively.

HRC Nr.	Test condition	HRC code	HRC group
1	Uncompressed 2D, content 1	2D1	
2	Uncompressed 2D, content 2	2D2	
3	Uncompressed 2D, content 3	2D3	
4	Uncompressed 2D, all content types	2D4	
5	Uncompressed anamorphic 2D,720p	2D5	2D
6	2D using the left view of 3D compressed at r04, 720p	2D6	
7	Compressed 2D at r02	2D7	
8	Uncompressed 3D, content1	3D1	
9	Uncompressed 3D, content2	D2	
10	Uncompressed 3D, content3	3D3	
11	Uncompressed 3D, all content types	3D4	
12	Uncompressed 3D, 720p SbS	3D5	
13	Simulated 3D (2D-to-3D conversion by geometrical distortion)	3D6	3D
14	Simulated 3D (uneven depth in vertical direction)	3D7	
15	Simulated 3D (temporal mismatch between left & right views)	3D8	
16	3D,720p SbS, compressed at r01	3Denc1	
17	3D,720p SbS, compressed at r02	3Denc2	
18	3D,720p SbS, compressed at r03	3Denc3	3Denc
19	3D,720p SbS, compressed at r04	3Denc4	
20	3D,720p SbS, compressed at r05	3Denc5	

2.3. Exp 3: Annoyance of crosstalk in S3D

A test was conducted where the crosstalk level in movie-like content was varied. A 3D projection system which could be used both with active and passive eyeglasses was used. The purpose of the test was to evaluate passive 3D projector system, but also to get some insight into the relationship between crosstalk and how visible and annoying the ghosting distortions are.

The set-up consisted of a DepthQ® HD3D projector from LightSpeed with a polarizing modulator from LC-Tec in front of the projector lens and a silver screen to project the sequences on for the passive eyeglasses. For the active eyeglasses the polarization modulator was removed.

The objective measurement of crosstalk was made at the centre of the screen. The measurement method adheres to ICDM standard[10]. The objective measured crosstalk from the projection system itself was about 0.3% for the system using active shutter eyeglasses and 2% for the system using passive polarized glasses (polarization modulator contributed less than 1%, the rest was due to other components in the system e.g. silver screen).

The procedure used for adding the crosstalk, was based on the measured system gamma function of the projector including the screen, which was found to be

$$L = 31.53 \cdot \left(\frac{Y}{255}\right)^{2.15}$$

where L is the luminance that was measure and Y the digital input Luma- or grey-values. The crosstalk is light leakage between the views, so the video Luma-values where then transformed into Luminance and the crosstalk were then added in this domain using the following equations

$$L_{left}^{crosstalk} = L_{left}^{original} + C \cdot L_{right}^{original}$$

$$L_{right}^{crosstalk} = L_{right}^{original} + C \cdot L_{left}^{original}$$

where C is the added crosstalk.

A subjective experiment with 26 naive test subjects was conducted in order to find out when a population of human observers starts to perceive the distortions and how the annoyance level could vary based on different amounts of crosstalk. The experiment consists of two main sessions: (a) passive projector system using passive polarized eyeglasses, and (b) active projector system using active shutter eyeglasses. The same test video set was shown to the subjects in both sessions.

The subjective experiment used Double Stimulus Impairment Scale (DSIS) as defined in ITU-R Rec. BT.500-13[7], using the five graded scale: imperceptible, perceptible but not annoying, slightly annoying, annoying and very annoying. Additional simulated crosstalk was added into the 3D test videos. Seven stereoscopic cinema contents were selected and processed in five simulated crosstalk levels (0%, 2%, 7%, 12 %, and 20%) plus the 2% system crosstalk for the passive system and plus 0.3% system crosstalk for the active system for the subjective experiment.

3. RESULTS

3.1. Exp 1: Multi-scale study for 3D video coding quality

A repeated measure ANOVA shows there are significant main effect between the scales, $F(2,48) = 1.39$ $p < 0.0001$, followed by a Tukey HSD post hoc test, together with the actual mean values, shows that the difference that comes from the ratings on discomfort scale are in general higher than the other two ($p = 0.0001$).

Video Quality and Presence are also highly correlated (correlation $= \sqrt{R^2} = 0.93$) as shown in Figure 1, but the regression line shows that the slope is less than one and the crossing of y-axis at x = 1 is about 2 rather than one, meaning that subjects have not giving as low values for presence when giving low video quality scores.

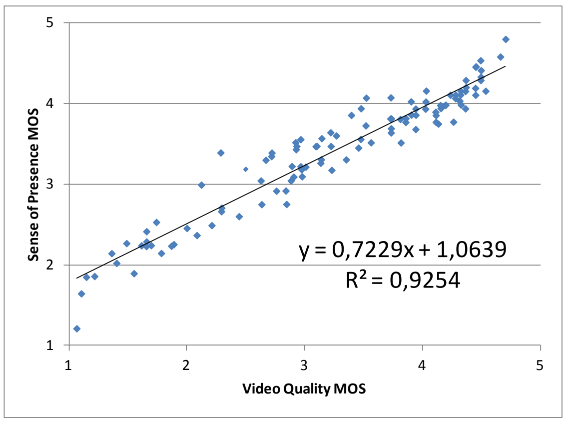

Figure 1: Relationship between Sense of Presence and Video Quality

Although video quality and discomfort were statistically significantly different, the correlation between the scale are very high (0.96), Figure 2. However, it can be noted that scatter plot shows a more bent behaviour than pure linear.

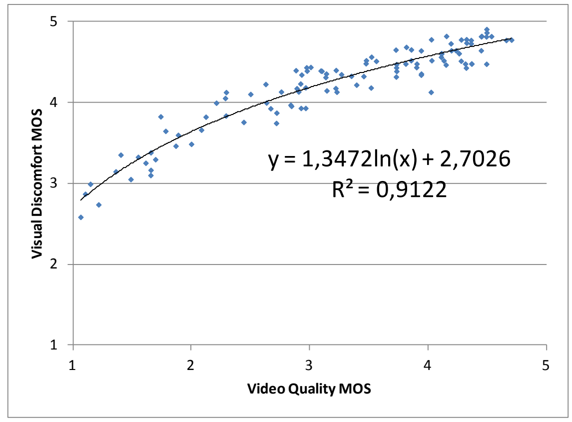

Figure 2: Relationship between Visual Discomfort and Video Quality

3.2. Exp 2: Multi-scale study of 3D video quality

A repeated measure ANOVA was performed, where the scales and the PVSs were treated as within-subjects factors and the different laboratories was a between-subjects factor, followed by a Tukey HSD post-hoc test which revealed that the Video Quality scales in the two labs were not statistically significantly different (p=0.8). Depth Naturalness and 3D Realism were not statistically significantly different on the 95% confidence level but close (p=0.1). The Depth Quantity was not significantly different from Depth Naturalness (p=0.8) or Video Quality (p=0.06), but different from 3D Realism (p=0.002). Visual Comfort was not statistically significantly different from the Video Quality scales (p=0.8 Lab1 and p=0.2 Lab2), but different towards the

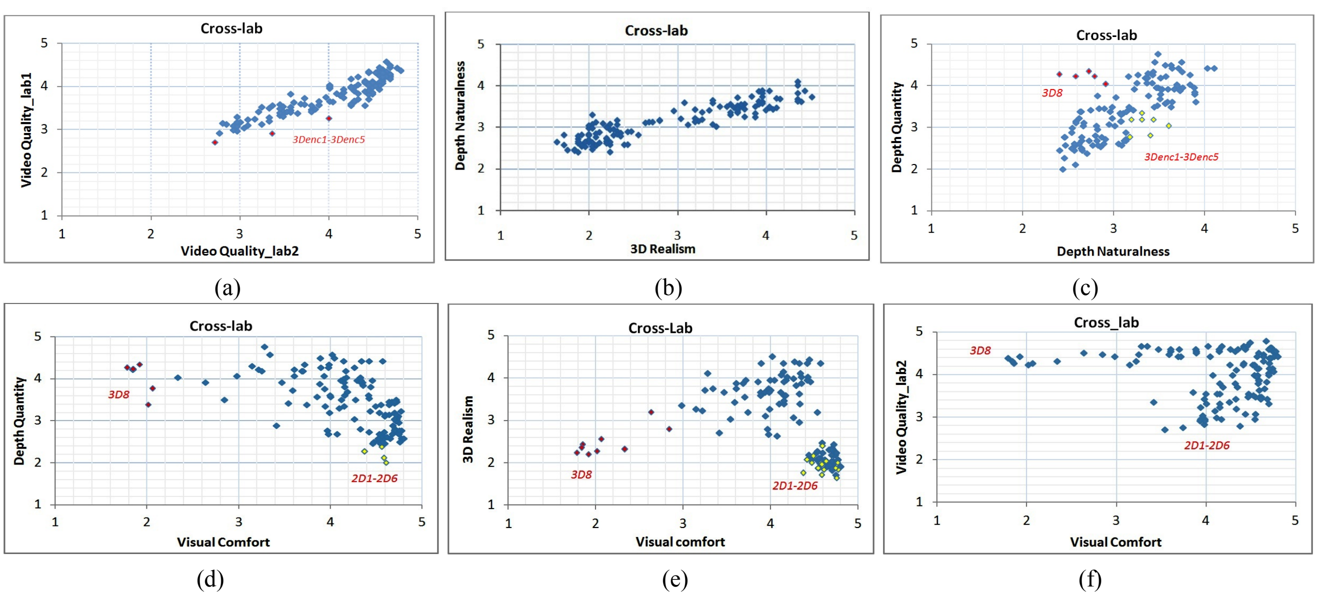

Figure 3: Cross-lab comparison

other scales (p=0.0001).

Table 2 reports the correlation between scales calculated for both PVS and HRC MOS results. Being more aggregated, correlation calculated based on HRC MOS is higher than the other in all pairs.

Table 2: Pearson correlation coefficient comparison

Scales	PCC (PVS MOS)	PCC (HRC MOS)
VQ1,VQ2	0.913	0.96
DN, 3DR	0.905	0.97
DN, DQ	0.68	0.90
VC, DQ	- 0.53	- 0.61
VC, 3DR	- 0.24	- 0.38

Figure 3 shows scatter plots of the MOS of the PVSs between different scales used in the two laboratories. According to Figure 3(a), the correlation between the video quality scales in the two laboratories was about 91% (see Table 2). Some PVSs had a bit different behaviour than the others which all belong to 3Denc HRC group. The Pearson correlation for DN-3DR is about 91% which is shown in Figure 3(b).. The behaviour of VC-DQ combination is shown in Figure 3(d) which indicates a reverse and not really high correlation between scales. The low correlation for the VC-3DR pair is shown in Figure 3(e). The VC-VQ2 pair had very low correlation, presented in Figure 3(f). One can also observe from Figure 3(e) and Figure 3(f) that the PVSs MOS are clustered, in two PVS groups with different correlations. For example Figure 3(e) reveals that all 2D PVSs (2D1-2D6) had similar VC and 3DR i.e. high VC but low 3DR. Figure 3(f) also reveals that some PVSs had similar VQ but different VC while other had correlated VQ-VC. More comparisons between the scales can be found in [2].

3.3. Exp 3: Annoyance of crosstalk in S3D

Both objective measured system crosstalk and simulated crosstalk are considered for the final evaluation of users' experience of perceived crosstalk. Figure 4 shows a linear relationship between the amount of overall crosstalk (sum of system crosstalk and simulated crosstalk) and users' Mean Opinion Scores (MOS) of perceptual crosstalk experience. The blue data points corresponds to the 3D projector system using active shutter eyeglasses, and the red points indicate the 3D projector system using polarization modulator and passive eyeglasses. The error bar indicates the 95% confidence interval of the mean over test subjects' individual voting on each watched video sequences. From the figure, we can see, it is necessary to control the overall crosstalk below 10% in order to keep the test subjects not annoyed (MOS > 3.5). We call this the acceptance level. The level when the test subjects starts to perceive the crosstalk distortion is about MOS=4.5 and this corresponds to about 3% crosstalk.

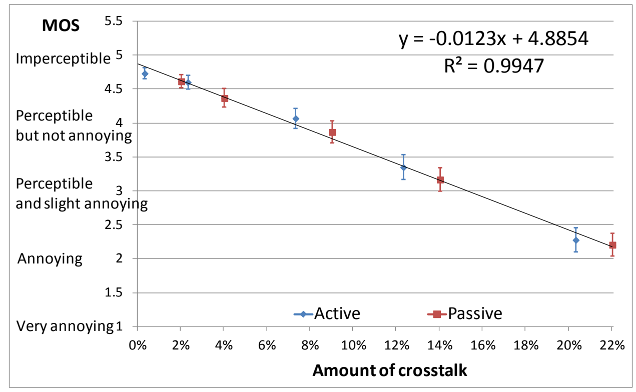

Figure 4: Relationship between amount crosstalk (in X axis) and users' Mean Opinion Scores (MOS) of perceptual crosstalk experience (in Y axis).

The perception of crosstalk distortions also has variations due to different video contents asFigure 5 shows. The X axis indicates the seven different video contents, the Y axis is an average of MOS across all HRC (crosstalk levels). The error bar indicates the 95% confidence interval of the mean over test subjects' individual voting on each watched video sequences. We can see, for SRC 3 and SRC 6, that people have had difficulties to distinguish different crosstalk distortions and therefore to a larger extent voted 4 or 5, also used much smaller part of the total scale.

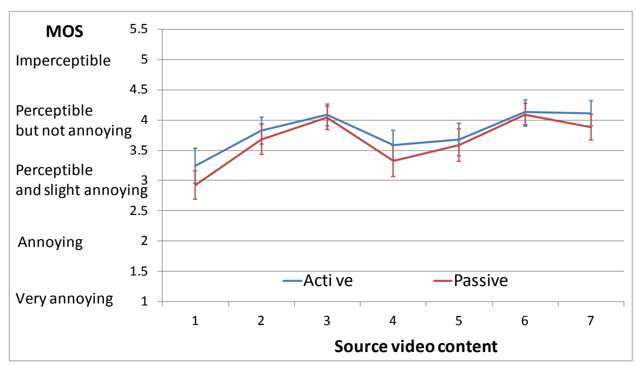

Figure 5: Relationship between source video content and user perceived distortions.

4. CONCLUSIONS

Three different experiments studied different aspects of QoE for S3D video.

Three scales: Visual Quality, Visual Discomfort, and Sense of Presence were studied in Exp 1. The results show that the different scales follow each other very well. This might indicate that one general video quality scale could be sufficient for evaluating S3D video experience, in some cases. That is mainly coding and spatial resolution reduction distortions. This is supported by Exp 2.

Exp 2 also confirmed that with a more complex mixture of degradations more than one scale should be used to capture the QoE in these cases.

The results show in Exp 3 that 10% can be considered as a crosstalk threshold for the test subjects not to be annoyed (MOS>3.5) by the distortions and thus acceptable. The distortions start to be perceived at about 3% crosstalk. The study found a linear relationship between perceived crosstalk and amount of crosstalk. The perceived crosstalk also varies largely depending on the video contents.

5. REFERENCES

[1] Brunnström, K., Ananth, I. V., Hedberg, C., Wang, K., Andrén, B., and Barkowsky, M., "Comparison between different rating scales for 3D TV", *SID Symposium Digest of Technical Papers Proc of SID Display Week 2013, May 21-24, 2013, Vanvouver, Canada,* paper 36.4 (2013)

[2] Kulyk, V., Tavakoli, S., Folkesson, M., Brunnström, K., Wang, K., and Garcia, N., "3D Video Quality Assessment with Multi-scale Subjective Method", *Proc of Fifth International Workshop on Quality of Multimedia Experience, QoMEX 2013, Klagen furt am Wörthersee, Austria,* paper 60 (2013)

[3] Wang, K., Hussein, M., Andrén, B., Brunnström, K., and Osterman, J., "Perception and annoyance of crosstalk in stereoscopic 3D projector systems", *Proc Stereoscopic Displays and Applications XXV 9011,* A. J. Woods, N. S. Holliman, and G. E. Favalora Eds., paper ? (2014)

[4] Lambooij, M., IJsselsteijn, W., Bouwhuis, D. G., and Heynderickx, I., "Evaluation of Stereoscopic Images: Beyond 2D Quality", *IEEE Transactions on Broadcasting* **57,** 432-444 (2011)

[5] Urvoy, M., Gutiérrez, J., Barkowsky, M., Cousseau, R., Koudota, Y., Ricordel, V., and Le Callet, P., "Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences", *Proc. Fourth International Workshop on Quality of Multimedia Experience, Yarra Valley,* (2012)

[6] ITU-T, "Subjective Video Quality Assessment Methods for Multimedia Applications", International Telecommunication Union, Telecommunication standardization sector, ITU-T Rec. P.910, (1999)

[7] ITU-R, "Methodology for the Subjective Assessment of the Quality of Television Pictures", International Telecommunication Union, Radiocommunication Sector, ITU-R Rec. BT.500-13, (2012)

[8] Jonsson, J. and Brunnström, K., "Getting Started With ArcVQWin", Acreo AB, acr022250, (2007)

[9] VQEG, "Report on the Validation of Video Quality Models for High Definition Video Content", Video Quality Experts Group (VQEG), www.vqeg.org, (2010)

[10] ICDM, "Information Display Measurements Standard (IDMS)", International Committee for Display Metrology (ICDM), Society for Information Display (SID), 1 (Version 1.03a), (6-1-2012)