Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11
3rd Meeting: Geneva, CH, 17–23 Jan. 2013

Document: JCT3V-C0202

Title: **Correlation analysis between MOS data collected on stereoscopic and autostereoscopic displays**

Status: Input Document

Purpose: Proposal

Author(s) or Contact(s): Tomasz Grajek Olgierd Stankiewicz, Krzysztof Wegner,
Tel: +48 61 665 3893
Email: kwegner@multimedia.edu.pl

Source: Poznan University of Technology, Poland

Abstract

This document tries to answer the question whether subjective evaluation can be performed only on one type of 3D monitor or various types of displays should be considered.

1 Introduction

Experiments conducted on responses of 3DVC Call for Proposals (CfP) [1] yielded huge amount of experimental data. In particular, 22 proposals in AVC and HEVC category were evaluated in two configurations: 2-view and 3-view. In 3-view configuration defined in CfP, three views along with estimated depth maps were fed into 3D Video Codec (Fig. 1). Decoded views and depth maps were then used in view synthesis module in order to produce a number of virtual views to be shown on stereo and autostereoscopic display. Exact positions of views produced and feed to the displays were specified in CfP and are presented in Table 1. Each generated stereo pair was centered over the middle coded view with baseline of 0.25 spacing between two neighboring inputs/transmitted views. On the other hand, autostereoscopic display shows 28 views spaced every 1/16th of baseline of two neighboring inputs/transmitted views, resulting in a much wider range of presented views (while each view is heavily decimated). Quality of synthesized material of each proposal was subjectively evaluated on both: stereo and autostereoscopic monitors during the formal evaluations of the proposals. A number of MOS data was collected for each proposal [2].

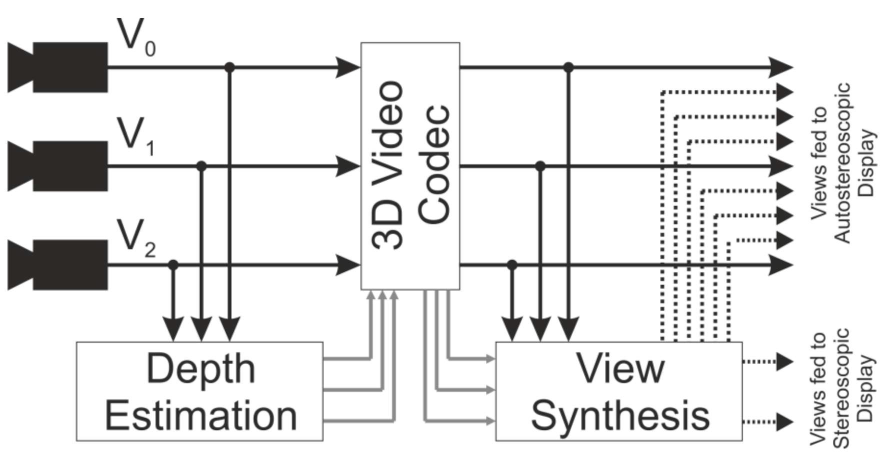


Figure 1 – Stereoscopic and auto-stereoscopic data for 3-view configuration

Table 1: Specification of views for stereoscopic and autostereoscopic displays

Seq. ID	Test Sequence	Test Class	3-view configuration		
			Input views	**Views for stereo pair**	**Views for autostereoscopic display**
S01	Poznan_Hall2	A	7-6-5	6.125-5.875	All 1/16 positions between views 7 and 5
S02	Poznan_Street		5-4-3	4.125-3.875	All 1/16 positions between views 5 and 3
S03	Undo_Dancer		1-5-9	4.5-5.5	All 1/4 positions between views 1 and 9
S04	GT_Fly		9-5-1	5.5-4.5	All 1/4 positions between views 9 and 1
S05	Kendo	C	1-3-5	2.75-3.25	All 1/8 positions between views 1 and 5
S06	Balloons		1-3-5	2.75-3.25	All 1/8 positions between views 1 and 5
S07	Lovebird1		4-6-8	5.75-6.25	All 1/12 positions between views 4 and 8
S08	Newspaper		2-4-6	3.75-4.25	All 1/12 positions between views 2 and 6

2 Problem definition

Subjective evaluation of the 3D video coding technology on both stereoscopic and autostereoscopic monitor is time and effort consuming. Therefore, a question arises whether evaluation of the 3D video on stereoscopic display is as good (gives the same statistically significant results) as evaluation of the same material on autostreoscopic display. It should be kept in mind that for both monitors we use the same decoded views and depth maps (see Fig. 1). However, for the stereoscopic monitor only stereo pair has to be rendered (based on those decoded views) and for the autostereoscopic monitor 28 views have to be synthesized (again based on the same decoded views).

3 Methodology

As mentioned before we want to check if there is any relation between subjective quality assessment obtained on both types of monitors. Therefore, we have analyzed correlation between results obtained on stereoscopic monitor and results obtained on autostereoscopic monitor. For each of sequence class we have fitted linear regression to the MOS obtained on both monitors

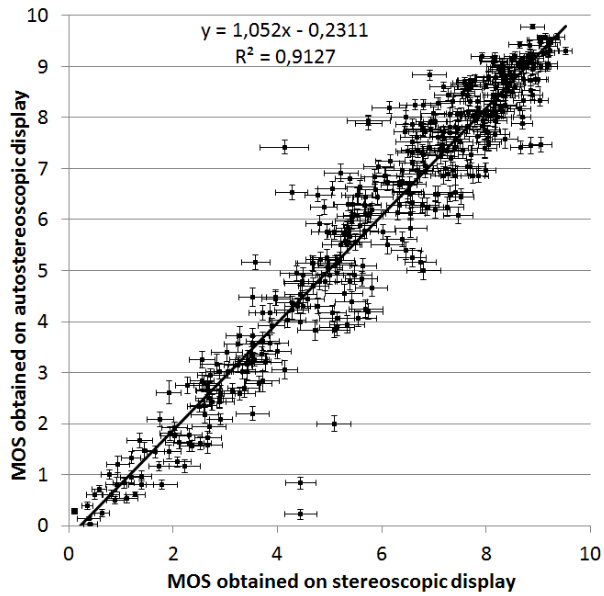
$$MOS_{autstereoscopic} = a \cdot MOS_{stereoscopic} + b \qquad (1)$$

and calculated Pearson correlation coefficient.

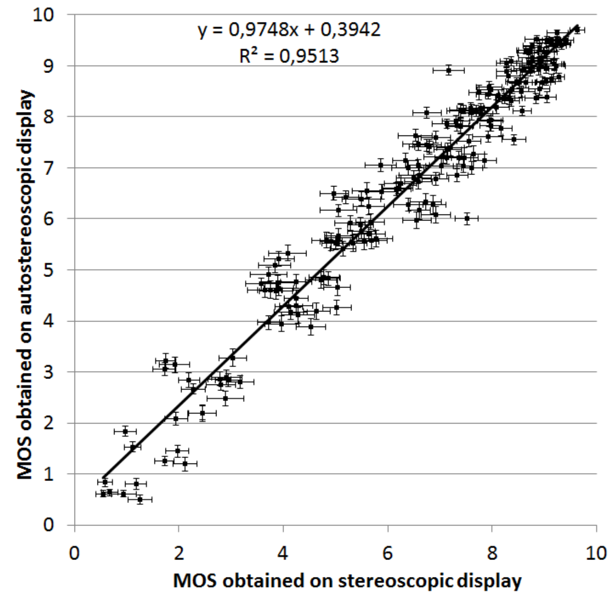
However, even high correlation is not enough to claim that results obtained on both monitors are the same. To make such a statement, we also need to know if the relation between scores of two randomly selected test points is the same on both monitors (see Fig. 3). In other words we want to check the following if the test point A outperform the point B on stereoscopic monitor, will it give the same result on autostereoscopic monitor. The answer for such a question will be given by Spearman rank order correlation coefficient. Therefore, we have ranked all the results and calculated the Spearman rank order correlation coefficient.

4 Results

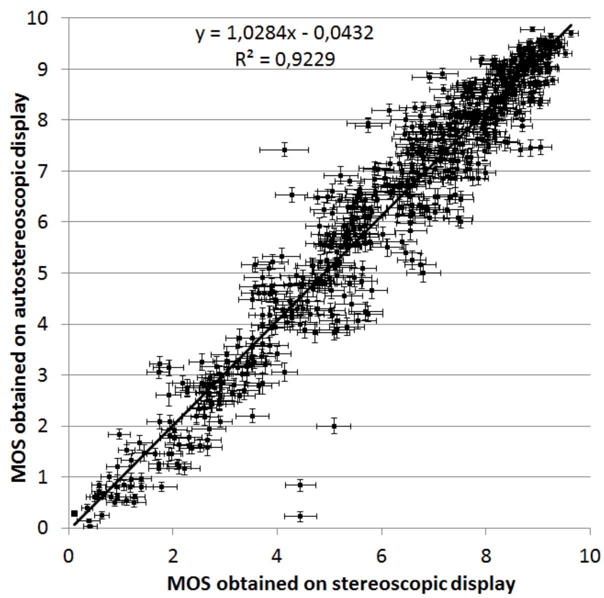
We have analyze MOS data obtained during MPEG evaluation of responses to CfP. Fig. 2 shows charts where on one axis we have MOS obtained on the stereoscopic monitor while on the other axis have MOS obtained on the autostereoscopic monitor for Class A (a), Class C (b) and Class A&C (c) sequences respectively. Additionally, 95% confidence intervals have been marked for all the test points. Table 2. presents correlation coefficient of MOS data collected on both types of Monitors.



a) Class A Full HD sequences



b) Class C XGA sequences



c) Class A and C

Figure 2. Linear regression fitted for data for stereoscopic and autostereoscopic monitors for class A (a), class C (b) and class A&C (c) sequences respectively.

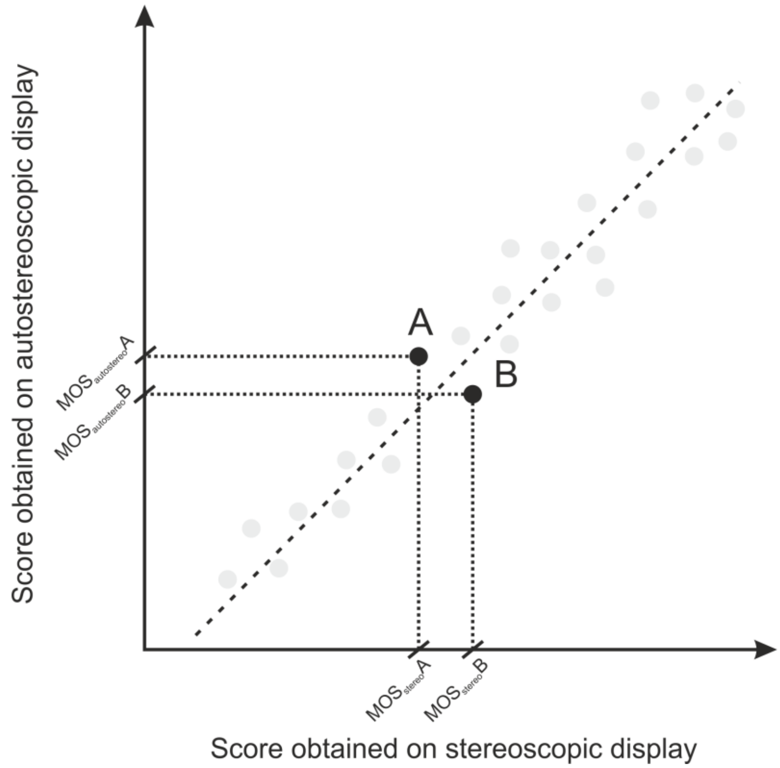


Figure 3. Linear correlation problem.

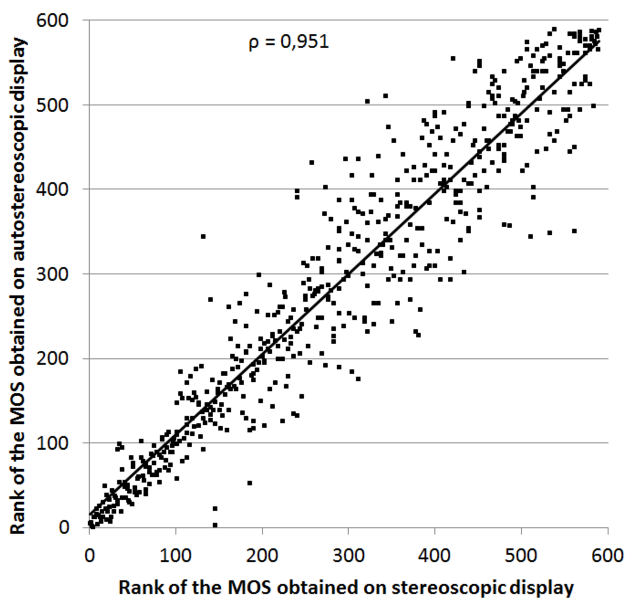


Figure 4. Spearman rank order correlation for Class A&C

Table. 2. Pearson correlation, Spearman rank order correlation and regression coefficients for analyzed data.

	Class A	Class C	Class A&C
Pearson correlation coefficient	0.955	0.975	0.961
Spearman rank order correlation coefficient	0.943	0.969	0.951
Regression coefficient	1.052	0.975	1.028

It can be noticed that results obtained on polarization monitor are well correlated with those obtained for the autostereoscopic one (Pearson correlation coefficient higher than 0.95 in all cases). Additionally, it has been proved that relation between scores of two randomly selected test points on both monitors is the same (Spearman rank order coefficient higher than 0.94). Moreover, scores obtained on autostereoscopic monitor are slightly better than these obtained on stereoscopic monitor, even though on autostereoscopic monitor we have lower spatial resolution of rendered views.

5 Conclusions

Our analysis showed that the results are independent from the display technology used. Therefore, it is worth considering to perform 3D subjective tests on one 3D monitor only (huge time and effort savings).

6 Acknowledgement

This work was supported by the public funds as a research project.

7 Patent rights declaration(s)

Poznań University of Technology may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).

8 References

[1] ISO/IEC JTC1/SC29/WG11, "Call Proposals on 3D Video Compression Technology", Doc. N12036, Geneva, Switzerland, March 2011.

[2] ISO/IEC JTC1/SC29/WG11, "Report of Subjective Test Results from the Call for Proposals on 3D Video Coding Technology", Doc. N12347, Geneva, Switzerland, November 2011.