4th International Workshop on Perceptual Quality of Systems (PQS 2013)
2-4 September 2013, Vienna, Austria

Toward Certification of 3D Video Quality of Experience through Inter - Laboratory Comparison

Dragana Đorđević[1], David Okolišan[1], Maja Pokrić[1], Jelena Kovač[1], Dragan Kukolj[2]

[1]RT-RK Institute for Computer Based Systems, 21000 Novi Sad, Serbia
[2]Department for Computer Engineering and Communication, University of Novi Sad, 21000 Novi Sad, Serbia

Email: {Dragana.Djordjevic;David.Okolisan;Maja.Pokric;Jelena.Kovac;Dragan.Kukolj}@rt-rk.com

Abstract

The fast development and increasing use of 3D video technology increases the need for standardization and certification processes when it comes to quality of video experience. Due to the specific conditions necessary for high-quality 3D video reproduction, an important role in evaluating the quality of 3D video plays the human experience factor. In order to find the best mechanisms for 3D video testing analysis of existing objective evaluation methods and proposal for certifiable subjective evaluations are required. This paper summarizes existing standards which are related to general inter-laboratory testing, examines the best practices for 3D video quality assessment from different laboratories, and based on gathered state-of-the art subjective measurements proposes a universal method which can be used as basis for certification procedure.

Index Terms: 3DTV, 3D video quality assessment, quality of experience, inter-laboratory testing, certification of QoE

1. Introduction

A well-recognized quality assessment testing procedure conducted by different laboratories, using identical video content and following similar methodologies and operations, can serve as an appropriate platform for demonstrating the certification procedure of the multimedia quality assessment. International standards, ISO/IEC 17025 [1] and ISO/IEC 17043 [2] provide description of the procedure for the cross-laboratory testing in general, and therefore they will be used as a framework for 3D inter-laboratory video quality testing. By following these standards, it is possible to carry out an evaluation of the factors that are crucial for generation of the test method. One of the most important aspects of inter-laboratory testing procedure is to provide accreditation of laboratories and equipment used in the process. ISO/IEC 17025 is used as the basis for accreditation of testing and calibration laboratories. ISO/IEC 17043[2] ensures that a laboratory operates according to the required standards and is capable of performing reliable and accurate proficiency testing on a continuous basis. It specifies general requirements for the competence of providers of proficiency testing schemes and for the development and operation of proficiency testing schemes.

Once you have provided all necessary parameters for laboratory accreditation, equipment and personnel competence, it is necessary to introduce the same method for all laboratories participating in the process of cross-laboratory testing. Participation in inter-laboratory comparisons schemes provides laboratories with an objective means of assessing and demonstrating the performance characteristics of the testing method. The method can be adopted as valid only if all results obtained by different laboratories, based on predefined criteria, are fulfilled.

It should be noted that inter-laboratory testing is necessary in order to assess the quality of 3D video that cover all aspects of objective and subjective measures of the Quality of Services or Quality of Experience to the end user. Therefore it is very important that all objective measures which are used in the quality evaluation are validated.

The rest of the document is structured as follows. Section 2 presents overview of inter-laboratory comparisons of 3D Video Quality of Experience and all significant aspects of 3D testing in cross validation procedure. Section 3 is the main part, comprising discussion about 3D testing procedure and describes the model given, which uses as a framework relevant standards, in an attempt to reach universal certification procedures for 3D video testing. The conclusions are written in Section 4.

2. Inter – Laboratory Testing of 3D Video Quality of Experience (QoE)

Inter-laboratory testing of 3D video quality, apart from having to adopt predefined general accredited procedures covers other aspects. Therefore, it is important to divide the entire process of testing in different phases, which also must include the objective and subjective assessment of video quality. Cross-laboratory3D video (3DV) testing based on ISO/IEC 17043 [2], can be defined as a process having the following phases: (i) selection of professional staff within an accredited institution, (ii) selection of the necessary equipment and environments that meet all the criteria of 3D video testing, (iii) design of proficiency testing schemes, (iv) design of the measurements statistical analysis, (v) introducing all participants to testing method , (vi) reporting of all results and finally (vii) communication with participants and exchange of information within confidentiality agreement. Certification process for the multimedia Quality of Experience (QoE) testing may include following four steps. The first step requires that the interested party (product manufacturer, service provider etc.) submits an application for certification, following that, the application information is analyzed by certification body. The second step includes assessment of the product or

10.21437/PQS.2013-12

service based on well defined criteria, by a laboratory previously accredited according to ISO/IEC 17025 [1]. In the third step, certification body reexamines whether all certification scheme requirements are fulfilled. If this is the case, product or service is certified. The fourth step involves periodic recertification according to certification scheme. Infrequently third parties, who must be accredited, can be involved in certification process. Third party has several obligations, firstly it is responsible for assessment, and secondly it must ensure that corresponding standards and rules are being applied in the assessment process. At last, it must provide appropriate certification mark or declaration of compliance. Accreditation body can permit this third party, if required standards and regulations are fulfilled, to provide testing and certification services. Confirmation check can be issued through external review, testing or audit, carried out by testing laboratory and certification body.

Figure 1 illustrates certification scenario in which Qualinet [18] group is involved. Qualinet comprises number of institutions, and is actively involved in multimedia quality of experience (QoE) domain.

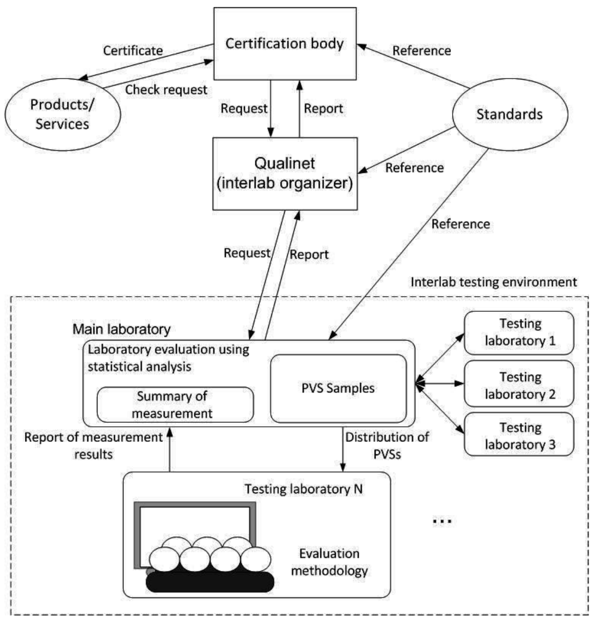

Figure 1: *An illustrative example of 3D inter-laboratory testing framework*

These institutions have ability and know-how to develop certification framework for inter-laboratory QoE which meet requirements of certification body. Qualinet may support the certification body with the testing framework necessary to organize and perform inter-laboratory quality assessment for products or services. Qualinet has built network of laboratories capable to run specific multimedia inter-laboratory testing assessment framework using certifiable procedures. They represent an example of third party, which can take part in recognized certification process that will ultimately lead to standardization of 3D video QoE testing.

3. Phases of 3D Video Quality of Experience Inter - laboratory Proficiency Testing

The inter-laboratory proficiency testing procedure can be further subdivided into phases which will lead towards easier standardization and certification. If we try to facilitate the process of testing, firstly it is necessary to summarize current state of standards. Table 1 summarizes the recommendation for parameters values based on relevant standards which must be respected in order to start certification process. Table 2 shows the comparison between 3D video related parameters used in practice for subjective quality assessment, drawn from several papers which described the procedures used to perform inter-laboratory testing. Both tables provide basis for the common framework for 3D video inter-laboratory testing certification process.

3.1. Personnel

Whenever you want to standardize and certificate some procedure it is necessary to consider personnel that will perform the entire process of testing. This is the first phase of proficiency testing and it provides required management and technical personnel. This means that minimum levels of personnel qualification and experience, authority for certain positions, educational and professional skills must be met by proficiency testing participants.

3.2. Equipment, Accommodation and Environment

This phase ensures that equipment is validated and maintained, and appropriate test area is prepared for proficiency testing scheme as defined by the ITU-R BT.500 [3]. In case of several laboratories participating in the testing, all laboratories must use the same evaluation setup which includes hardware, software and working ambient. For 3D proficiency testing scheme some relevant parameters are given in Table 1. Firstly, display characteristics such as peak luminance (70-250 cd/m^2), contrast ratio ($\leq$0.02) and monitor resolution are given. The values in Table 1 are recommended by ITU-R standards, but in practice there are limitations that affect parameters. The peak luminance is parameter that must be adjusted according to the room illumination, while the contrast ratio is affected by crosstalk [11]. Parameters like background and room illumination are critical because they depend on the minimum distance between the display and the background. Viewing distance and position, on the other hand, depend on the resolution and the size of display. Depth rendering is not specified by any standard but many papers [11][12] describe this parameter as one of the most important affecting the quality of 3D video. In [11] authors give detailed definition of depth rendering and concluded that it depends on viewing distance, content disparity and display properties.

Table 1: *Recommended parameter values by ITU-R standards*

ISO/IEC 17043	Parameters	Recommendation	ITU-R
Equipment, accommodation and environment	Peak luminance	*70-250 cd/m²*	[3][7][8]
	Contrast ratio	*≤0.02*	[7][8]
	Background illumination	*≈0.15*	[7]
	Room illumination	*Low*	[3][4][7][8]
	Monitor resolution	*1280×720 SDTV 4.8H and 1920x1080 HDTV 3.1H*	[7][8]
	Viewing distance	*Table 4 [8]*	[4][8]
	Viewing position	*For 1280 × 720 : 16:9 aspect ratio, 1 21° angle; For 1920 × 1 080: 16:9 aspect ratio, 1 31° angle*	[7][8]
	Depth rendering	*Not specified*	-
Design of proficiency testing schemes and choice of method or procedure	Signal format	*Figure 1* in *[9]*	[9]
	Signal format conversion	*Not specified*	-
	Signal content complexity	*2D video complexity: Spatial Perceptual Information (SI), Temporal Perceptual Information (TI)*	[10]
	Visual comfort limits	*Screen parallax: 1% for crossed/negative disparities and 2% for uncrossed/positive disparities; Depth of field: between ±0.2D and ±0.3D; Retinal disparity: ±1° of visual angle*	[8]
	Number of observers	*30*	[8]
	Session duration	*~20-40 minutes intermixed with breaks*	[3][8]
Operation of proficiency testing schemes	Training and introduction	*Five "dummy presentations" to stabilize the observers' opinion*	[3]
	Grading scale	*Table 1 [8]*	[3][8]
	Test protocol features	*Video segment, Test condition, Session, Test presentation, Voting period, Segments of votes*	[3]
	Evaluation method	*SS, DSCQ, SC, SSCQE, DSIS*	[3][8]
Data analysis/evaluation of proficiency testing scheme results	Vision screening	*Snellen charts - visual acuity; Ishihara plates - color; Randot, Stereo Fly, Frisby tests - stereoscopic vision (more tests in Table 5 [8])*	[6][8]
	Viewers' rejection	*Systematic shifts, local inversions*	[3][8]
	Statistical analysis	*Statistical analysis of each separate video segment (VS); Statistical analysis of each separate test conditions (TC); Overall statistical analysis of all the pairs VS/TC*	[3][8]
	Multi-step analysis	Means and standard deviations: for each observer vote, for each SOV; Statistical distribution of the means calculated at the previous step; The global annoyance characteristic	[3][8]
Reports	Presentation	*Charts, table, scale*	[3][6]

We also analyzed published reports from scientific experiments performed by several research laboratories as examples of good practice. For example, laboratory of IRCCyN (University of Nant, France) and Acreo AB (Sweden) used the same 23" display (120Hz, resolution 1920x1080p) for the experimental setup [15] [14] along with a pair of active shutter glasses. The display was positioned far enough from the wall to avoid any interference between the displayed 3D content and the real world. The viewing distance was set to be 3 times of the display height, and the same value was used in the VQEG HDTV test plan [17]. At IRCCyN, backlighting was used to adjust the lab luminance, while the luminance level of the reflection from the gray wall behind the screen was set to 50cd/m², which corresponded to 15% of the peak luminance of the display after passing the shutter glasses when they were activated. At Acreo, the room luminance was set to 20 lux, which is close to darkness, in order to avoid reflections from other objects.

In another research [16], conducted by laboratories of IRCCyN (University of Nant, France) and Polytechnic University from Madrid, observers were seated in a standardized room. A 46" stereoscopic display with shutter glasses was used to display the sequences. Screen brightness was set to 180 cd/m2, resulting in 56 cd/m² after glasses luminosity attenuation. Room illumination was calibrated as well, and set to 54 cd/m² behind the screen, which resulted in 8 cd/m² after glasses attenuation, hence 15% of the perceived screen brightness. Finally, viewing distance was set to three times the height of the screen (i.e. 172 cm), similarly to previous research.

Laboratories Acreo, EPFL, UBC and NTNU have jointly conducted the inter-laboratory testing of 3D video QoE [13], wherein they used the same equipment and environment setup. They used 46" polarized stereoscopic monitor and 52" auto-stereoscopic monitor with native resolutions of 1920x1080 pixels and the same test video material. The hardware and software used in all laboratories was designed and tested to meet specified requirements before actual evaluations took place. The laboratory setup was controlled in order to ensure the reproducibility and validity of results.

3.3. Design of Proficiency Testing Scheme

When modeling a scheme for inter-laboratory testing it is necessary to pay attention to the planning, preparation, statistical analysis, the selection of the test material and documentation. The step of planning identifies and documents processes that directly affect the quality of proficiency testing programs. In the preparation step, it is necessary to establish and implement procedures to ensure appropriate acquisition, preparation, handling and storage of the test material. The design should take into account statistical significance of the results, therefore the following has to be taken into account: the accuracy, the minimum number of participants, the number of proficiency test items and number of repeated tests, procedures used to establish the standard deviation, the procedures for the evaluation of values excluded from statistical analysis (i.e. outliers). Finally, there is a need for documenting the whole test procedure and obtained results.

Table 1 gives the parameters that should be taken into account in this phase of testing process. First of all video material should be selected. This is not easy task because video material has its own characteristics such as signal format, signal format conversion, complexity, and length of the material. Various signal formats can be found in standards and literature, such as: conventional stereo video, 2D-plus-depth, MVC (multi-view), MVD (multi-view plus depth), LDV (layer depth video) DES (depth-enhanced stereo) [9][11]. Signal content complexity is standardized for 2D video (Perceptual Spatial Information SI, Perceptual Temporal Information TI), but for 3D video Depth Perceptual Information DI [12] has to be taken into account. Also, it is necessary to define a number of observers who will take part in testing phase.

Table 2: *Parameters used in3D subjective inter-lab testing schemes*

Parameter	[13]	[14]	[15]	[16]
Sequence	*Stereoscopic and auto-stereoscopic*	*Stereoscopic*	*Stereoscopic*	*Stereoscopic*
Down-sampling	*Only color sampling is specified*	*Spatial and temporal*	*Spatial and temporal*	*Type is not specified*
Video encoding	*H.264/AVC, HEVC*	*H.264/AVC, JM17.0, JMVC7.1*	*H.264/AVC, JM17.0, JMVC7.1*	*H.264/AVC, JM18.2, JPEG2000*
Video decoding	*Not enough described*	*JM15.1*	*Directly*	*Not specified*
Quantization parameter (QP)	*Not specified*	*26-44 (step 6)*	*26-44 (step 6)*	*32-44*
Up-sampling	*Only color sampling is specified*	*To original temporal frame rate*	*To original temporal frame rate*	*Type not specified*
Depth	*Depth map described*	*Not clearly shown*	*Not clearly shown*	*Depth map described detailed*
Frame rate	*25Hz, 30Hz*	*24Hz, 25Hz*	*24Hz, 25Hz*	*25Hz*
Resolution	*1920x1088, 1920x768*	*1920x1080, 1280x720, 1440x1080, 1024x576*	*1920x1080, 1280x720, 1440x1080, 1024x576*	*1920x1080*

Recommendation [8] specifies 30 observers, while earlier standards [3][4] recommend 15 observers. However, this parameter greatly depends on predefined sensitivity and reliability required by the experiment [12].

In practice, there are more parameters that are necessary to consider during the preparation of data. Table 2 provides an overview of parameters that are not yet clearly standardized but their values greatly affect perceived subjective quality of 3D video material. Table 2 gives an overview of only publically available inter-laboratory studies.

In [13] auto-stereoscopic and stereoscopic *video sequences* were used, while in other studies emphasis was on stereoscopic sequences. In [14] [15] spatial and temporal *down-sampling* was performed, while [12] only mention color sampling (with 8 bits precision per pixel). Research [16] describes HRCs feature sequences down-sampling with factor 4, and later up-sampling for displaying. In this case losses in resolution were taken into account.

All studies used H.264/AVC *video encoding*, with various additions. For example in [13] HEVC, random access encoder and MVC encoder JMVC 8.3.1 were used. In [14][15] JM 17.0 and JMVC 7.1 beside H.264/AVC were used. JM 18.2 and JPEG 2000 still image coder (Kakadu software, v7.0) was used in [16]. Decoded sequences in [14][15] were upsampled to their original temporal frame rate and the full HD resolution which is the same as the native display resolution of the 3D screens used in the subjective experiments. In addition, in [14], transmission network was assumed and bits were decoded using JM15.1, while in [15] transmission step was skipped and bits were decoded directly. Value of quantization parameter was not provided in [13], but in [14][15] was set to 26 to 44 with step size 6. Incrementing the QP by six, doubles the quantization step size of the linear quantizer for the DCT coefficients in the H.264 encoder. This also approximately halves the bitrate. In [16] QP varies between 32 and 44.

Utilization of *depth map* was described in [13] and [16], while in [14][15] depth was mentioned as one of key parameters in subjective video quality evaluation. Frame rate in all studies was 25 Hz, with additional frame rates of 24 Hz and 30 Hz [14][15].

3.4. Operation of Proficiency Testing Schemes

According to [2], all testing participants should be given the option to choose the test method and procedures. However, the proficiency testing provider may instruct participants to use a specified method in accordance with the design of the proficiency testing scheme. Operation of proficiency testing includes introducing all participants to testing method, testing material, and testing process itself. Looking at the results of different experiments in cross-laboratory testing it can be noticed that there are specific steps and methods that were used in most subjective assessment experiments of 3DTV quality.

Phase “Operation of proficiency testing” is presented in Table 1 with the standardized parameters: training and introduction, grading scale, test protocol features and evaluation method. Analyzing results from different experiments in cross-laboratory testing it can be noticed that there are specific steps and methods that were mostly used in subjective assessment of 3DTV quality. In references [15][14][16] is described the “absolute category rating with hidden reference” (ACR-HR), subjective method which allows estimation of the user experience when it comes to assessing the stereoscopic video quality, as well as comparison of different coding and transmission scenarios and performances of video compression standards. Subjective tests in different laboratories with the ACR method gave quality judgments based on two panels of observers. In this experiment, the processed video sequences (PVSs) were presented in random order and participants rated PVS independently on the ACR category scale which is five-point quality scale defined by [10]. Each time prior to running the experiment participants must pass training. In addition to visual explanations the most laboratories provide a written version of the instructions as described in [12]. It should be mentioned that the training and test is usually implemented using a specially designed graphical user interface (GUI). Some laboratories used the DSCQS method for the training session [14], or DSIS evaluation methodology [12].

3.5. Analysis of Proficiency Testing Scheme Results

After training, and testing phase, the test, results received from participants must be recorded and analyzed by appropriate methods. Procedures must be established and implemented to check the validity of data entry, data transfer, statistical analysis, and reporting [2]. At this stage, in order to be able to perform the inter-laboratory comparison, the results of participating laboratories are collected and correlated in systematic manner. Inter-laboratory subjective video quality assessment is most commonly represented by Mean Opinion Score (MOS) calculated from scores given by number of different viewers for a particular standardized test set-up. The first step of data analysis procedure is checking the validity of the results and excluding the outliers from the data set. Then, on the valid data set suitable inter-laboratory analysis can be performed.

Table 3 summarizes proficiency testing scheme parameters which were used in practical testing. Summarized papers represent the good practice examples and many of the parameters can be adopted in certification of unique scheme for testing of 3DV. During the procedures described in [13] [14] [15], the observers were screened and their visual capabilities measured using a *Snellen Chart*, stereo vision and depth perception acuity using a *Randot Stereo test*, and color blindness using a Ishihara plates. After the experiment all votes were checked ([3],[14]) and some of them were rejected. In [16] diverse information on the subjects was included to build a dataset of observers. Observers were asked how experienced they are with stereoscopic 3D contents, and empirically attributed a note ranging from “never experienced 3D” to “everyday use”. In addition, ocular tests can be performed ([13] [14] [15]) with additional eye dominance - Porta test. With respect to [3], all observers were labeled as outliers. With respect to [10], all observers were accepted, although the correlation of some observer for all PVSs is only 0.75, which is the rejection threshold.

Table 3: *Evaluation parameters of proficiency testing scheme*

Parameter	[13]	[14]	[15]	[16]
Number of labs	*12*	*2*	*2*	*2*
Voting interface	*GUI on screen, paper scoring sheets*	*GUI on separated screens*	*GUI on separated screens*	*GUI on screen, results are recorded*
Voting methodology	*DSIS*	*ACR-HR, DSCQS*	*ACR-HR*	*ACR-HR*
Stimulus between presentation	*Grey screen*	*Presentation followed immediately after an observer validate vote*	*Presentation followed immediately after an observer validate vote*	*Grey screen*
Session duration	*7 seconds for each sequence*	*16 seconds per sequence*	*16 second per sequence*	*16 seconds + 13 seconds*
Vision test	*Snellen chart, Randot test, Ishihara chart*	*Snellen chart, Randot test, Ishihara chart*	*Snellen chart, Randot test, Ishihara chart*	*Snellen chart, Randot test, Ishihara chart, Porta test*
Experience test	-	*Observers’ experience 3D video quality*	*Observers’ experience 3D video quality*	*Observers’ experience 3D video quality*
Rejection of participants	-	*7 observers*	*7 observers*	*1 observer*
Statistical method	*ANOVA, Spearman and Pearson correlation, Student t-test*	*ANOVA, Fisher-Snedecor distribution*	*ANOVA, Fisher-Snedecor distribution*	-
Additional analysis	-	*Rate-distortion analysis*	*Bit rate/MOS analysis, Gender analysis*	*Comparison of MOS for original and impaired sequence*

Analysis of data validity can be done using different criteria. The goal of this analysis is to assess credibility of each subject which contributed to laboratory test results. As measures of subject’s evaluation credibility can be used: assessor sensitivity, assessor reproducibility and assessor agreement. Assessor sensitivity stands for measuring each assessor's ability to identify differences between video sequences. Evaluation of each single subject to reproduce their scores is called assessor reproducibility. For evaluation purposes if the subjects have rated the test data similarly we could estimate assessor agreement. Comparison of results could be performed using various statistical methods such as *Spearman* and *Pearson correlation*, *ANOVA*, *Brown-Forsythe test*, *Student’s t-test*, *Welch’s test*, *Kruskal-Wallis* test etc. In [13] 3DV inter-laboratory comparison experiment, Spearman and Pearson, ANOVA and Student’s t-test were used for results cross-correlation between different laboratories. The Pearson linear coefficient measured the distribution of the MOS points given for different viewers from selected laboratories around the linear trend, while the Spearman coefficient measured the monotonicity of MOS between different laboratories. ANOVA was used between group variables and the different PVS as within group variables. In addition, a Student’s t-test was applied to each pair of PVS from the different laboratories, identifying the significantly different PVS.

When processing the data obtained experimentally there can be more independent analysis such as cross-lab comparison, gender analysis, rate-distortion analysis and analysis of observers’ experience of 3D video quality as is the case in [15].

For cross-lab comparison in [15] Analysis of Variance (ANOVA) was performed with the laboratories as one between factor and data sets as within factors. They concluded that there was difference between laboratories which can be reduced by linear rescaling. The analysis of gender showed that there was a significant interaction between the genders and the HRCs part of data (Hypothetical Reference Circuits) because females voted more positively on high quality videos than males, but also decreased their quality ratings faster than males.

Based on the analysis of bitrate gain, which indicates the amount of bitrate that can be saved when the MOS remains constant and “Quality gain” factor for a given bitrate [15] it was shown that for HD 3DTV transmission system, reduction of the resolution by a factor of four before the video encoding will result in a better quality.

Experiment from [14], together with analysis from [15], also contains cross-experiment comparison analysis. The first six HRCs were used in the coding efficiency and the packet loss data set. After obtaining the combined data for the cross-lab coding efficiency experiment, these six HRCs were extracted and compared to ones obtained in the packet loss experiment. During the analysis of data [16] emphasis was placed on the variability of the characteristics of the observers, observer’s votes’ distribution and MOS distribution for original and impaired sequences. It is shown that proposed degradations, impairments and post-processing data spread over the whole ACR scale. In addition, MOSs obtained for all HRCs are trustworthy as corresponding confidence intervals are narrow. From this information, one may compare the influence of standard video coder’s artifacts and those of wavelet based coders, on performance of trained MOS estimator.

3.6. Reports

The report phase describes how test results should be reported. This is an important deliverable for ease of comparison between the tests and results obtained by different laboratories. The rule [2] is that all analyzed tests have to be documented. Report must contain numerical results and a detailed description of the procedure that was used. Also, all results should be displayed graphically to help understanding.

Depending on affinities and needs when analyzing the data, reports and graphical presentation of results may differ greatly. However, if we consider [13][14][15] it can be seen that the scatter plot, presenting the difference between laboratories, is representation of choice. Most reports also include a graphical representation of MOS versus EMOS across stereoscopic video sequences or HRC of different laboratories [14][15], comparison of 3D and 2D reference sequences and MOS versus EMOS for each laboratory. Additionally, the reports could present many comparisons of coding artifacts, spatial and temporal downsampling, 3D EMOS for H.264 coding etc. Variations of these graphic presentations can be very diverse.

3.7. Communication with Participants and Confidentiality

Communication with participants and confidentiality provide means of controlled exchange of detailed information about the proficiency testing scheme. To increase efficiency of information exchange it has to be predefined what can be disclosed by each party.

4. Conclusions

In order to achieve precision in inter-laboratory testing and improve the quality of measurement of 3D video experience, that includes subjective and objective assessment, it is essential to establish a uniform testing scheme. The goal of this paper is to propose for 3D video quality assessment the most acceptable testing framework based on existing standards and state-of-the-art methods in hope that this will lead to certifiable process of inter-laboratory testing. Establishing a unique method of testing is not an easy task, because it is necessary to analyze all parameters and conditions which may affects 3D quality. As presented, many things greatly influence the quality of experience of 3D multimedia contents, such as: the environment, equipment, light, distance and the participants in testing. Currently, many laboratories are collaborating towards common goal of establishing unified and widely acceptable inter-laboratory testing procedures for validation of products and services related to 3D multimedia content. This paper proposes a framework which leads to certification of 3D video testing.

5. Acknowledgements

This work was partially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia under Grant TR-32034, by the Secretary of Science and Technology Development of the Province of Vojvodina under Grant 114-451-2636/2012-01 and Qualinet-COST Action IC1003.

6. References

[1] ISO/IEC 17025:2005, *General requirements for the competency and testing of calibration laboratories*

[2] ISO/IEC 17043:2010, *Conformity assessment — General requirements for proficiency testing*

[3] ITU-R BT.500-13 (2012, January), *Methodology for the subjective assessment of the quality of television pictures*, International Telecommunication Union

[4] ITU-R BT.710-4 (1998, November), *Subjective assessment methods for image quality in high-definition television*, International Telecommunication Union

[5] ITU-R BT.1788 (2007), *Methodology for the subjective assessment of video quality in multimedia applications*, International Telecommunication Union

[6] ITU-R BT.1438 (2000), *Subjective assessment of stereoscopic television pictures*, International Telecommunication Union

[7] ITU-R BT.2022 (2012, August), *General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays*, International Telecommunication Union

[8] ITU-R BT.2021 (2012, August), *Subjective methods for the assessment of stereoscopic 3DTV systems,* International Telecommunication Union

[9] ITU-R BT.2160 (2012, June), *Features of three-dimensional television video systems for broadcasting*, International Telecommunication Union, International Telecommunication Union

[10] ITU-R P.910 (1999, September), *Subjective video quality assessment methods for multimedia applications*

[11] Chen W., Fournier J., Barkowsky M., Le Callet P., (2010, March 12), "*New requirements of subjective video quality assessment methodologies for 3DTV*", VPQM 2010, Scottsdale, Arizona, USA

[12] Chen W., Fournier J., Barkowsky M., Le Callet P., (2013, March), *Methodologies for Assessing 3D QoE: Standards and Explorative Studies*, ZTE COMMUNICATIONS Vol.11 No.1

[13] Perkis A., You J., Xing L.,Ebrahimi T., De Simone F., Rerabek M., Nasiopoulos P., Mai Z., Pourazad M. T., Brunnström K., Wang K., Andrén B., (2012, January 19), "*Towards certification of 3D video quality assessment",* 6th Int. Workshop VPQM 2012, Scottsdale, Arizona, USA.

[14] Wang K., Barkowsky M., Brunnström K., Sjöström M. , Cousseau R., Le Callet P.,(2012, August 20) *Perceived 3D TV Transmission Quality Assessment: Multi-Laboratory Results Using Absolute Category Rating on Quality of Experience Scale,* HAL

[15] Barkowsky M., Wang K., Le Callet P., Cousseau R., Brunnström K., Olsson R., Sjöström M., *Subjective evaluation of HDTV stereoscopic videos in IPTV scenarios using absolute category rating,* HAL 24.August 2011

[16] Urvoy M. , Barkowsky M., Cousseau R., Koudota Y., Ricordel V., Le Callet P., Gutiérrez J., García N., *NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences,* QoMEX, 2012 Fourth Int. Workshop on 5-7 July 2012

[17] Cermak, G., L. Thorpe and M. Pinson; *Test Plan for Evaluation of Video Quality Models for Use with High Definition TV Content*, Video Quality Experts Group (VQEG), (2009).

[18] Qualinet Online: http://www.qualinet.eu/