Subjective evaluation of HDTV stereoscopic videos in IPTV scenarios using absolute category rating

Kun Wang, Marcus Barkowsky, Romain Cousseau, Kjell Brunnström, Roger Olsson, Patrick Le Callet, M. Sjöström

HAL Id: hal-00616712
https://hal.science/hal-00616712v2

Submitted on 24 Aug 2011

Subjective evaluation of HDTV stereoscopic videos in IPTV scenarios using absolute category rating

K. Wang[a,c*], M. Barkowsky[b], R. Cousseau[b], K. Brunnström[a], R. Olsson[c], P. Le Callet[b], M.Sjöström[c]

[a]Dept. of NetLab: IPTV, Video and Display Quality, Acreo AB, Sweden
[b]Dept. of Image and Video Communication, IRCCyN, University of Nantes,France
[c]Dept. of Information Technology and Media (ITM), Mid Sweden University, Sweden

ABSTRACT

Broadcasting of high definition (HD) stereobased 3D (S3D) TV are planned, or has already begun, in Europe, the US, and Japan. Specific data processing operations such as compression and temporal and spatial resampling are commonly used tools for saving network bandwidth when IPTV is the distribution form, as this results in more efficient recording and transmission of 3DTV signals, however at the same time it inevitably brings quality degradations to the processed video. This paper investigated observers quality judgments of state of the art video coding schemes (simulcast H.264/AVC or H.264/MVC), with or without added temporal and spatial resolution reduction of S3D videos, by subjective experiments using the Absolute Category Rating method (ACR) method. The results showed that a certain spatial resolution reduction working together with high quality video compressing was the most bandwidth efficient way of processing video data when the required video quality is to be judged as "good" quality. As the subjective experiment was performed in two different laboratories in two different countries in parallel, a detailed analysis of the interlab differences was performed.

Keywords: ACR, QoE, 3DTV, Simulcast coding, Multiview coding, spatial resampling, temporal resampling

1. INTRODUCTION

3D videos are riding their success from cinema to home entertainment markets such as TV, DVD, Blu-ray, video games, etc. As those popular 3D films go through the box office, a huge profit could be foreseen from bringing the 3-D experience to the home. A number of techniques have been invented and are in most cases still under fast development for representing and presenting 3D videos, e.g. multi-view, 2D video plus depth, volumetric, however, the side-by-side S3D is the most common and mature technology that has been widely used in the movie industry and today's 3DTV broadcasting for its simplicity and compatibility.

S3D videos present viewers two similar images with slight visual disparity, i.e. two perspectives of the same object. With the help of the glasses each eye of the viewers will only see one of the pictures. The brain will then put corresponding points and objects together in the two images and create a sense of 3D depth through disparity. In the home environment, the channels currently used for IPTV seem most appropriate for S3D services in HDTV quality. However the transmission bitrates has a strong impact on the delivered video quality, the bandwidth requirements still a challenge for both broadcast and streaming.

Video compress encoding is one of the most common techniques for saving the transmission bandwidth. The encoding technology for 2D videos has progressed with the development of H.264/AVC coding scheme. Today, most of common 3D coding methods are based on H.264[1], e.g. H.264 simulcast coding (independent bitstream coding) and multi-view coding[2] (MVC, dependent bitstream coding). For stereoscopic videos, simulcast coding encodes the left and right views separately with H.264/AVC in the very same way as it was handled in normal 2D videos. The advantage of H.264/AVC is that it can be achieved by current standards and existing hardware, hence it allows the broadcaster to use most of its 2D infrastructure even for 3D. However as the images of the different views are highly correlated a lot of information between the two views is redundant. MVC is one of the codec that uses this redundancy to improve coding efficiency by introducing inter-view prediction, where images are not only predicted from temporally neighboring images but also

---

from corresponding images in adjacent views. Besides encoding techniques, specific data processing operations such as temporal and spatial resampling are tools that are also frequently used for more efficient recording and transmission of video signals. However when it comes to 3D videos, the question is still open which ones of the processing steps are efficient for saving bitrate as well as their impacts on perceived 3D visual quality.

To save the transmission bitrate of 3D videos, some researchers have focused on asymmetric video processing. Saygili, G.(2009)[3] employed asymmetric coding where one view was encoded by H.264/AVC and the other using scalable video coding. Scalable multiview coding (SMVC) with temporal and spatial scalability had been studied in Ozbek,N.'s (2007) paper[4]. In Stelmach, L. (2000)[5] the mixed resolution or frame rate reduction in one of the stereo views were explored.

The evaluation of the perceived video quality is of highest importance from the end user's point of view. In the 2D video case, it has been well studied in many publications, e.g. Pinson, M (2010)[6], Barkowsky, M. (2010)[7], and several different methods for measuring the subjective quality have been developed and tested.In the 3D case, the most common approach is to use the Double-Stimulus Continuous Quality-Scale (DSCQS) method as specified in ITU-R BT.500[8] on stereoscopic displays[3, 4, 5] for 3D presentation. However, the DSCQS method is time consuming since each video sequence is played at least twice, which limits the number of test sequences that can be evaluated during the subjective test. As an alternative, the Absolute Category Rating (ACR) as specified in ITU-T P.910[9] may be used which only plays each video once, therefore it is capable of collecting more votes from subjects. This allows to test a broader range of qualities for example on varies scenarios of high definition (HD) S3D videos. There are some other methods that have been used for the 3D subjective quality evaluation as well, such as Single Stimulus Multimedia (SSMM) stated in Merkle, P. (2007)[10] which is a modified version of the Single Stimulus Impairment Scale where the same video was shown twice to viewers before rating, but there is no unimpaired reference videos in the test sequence set. Strohmeier (2010)[11] introduced a mixed method called Open Profiling of Quality (OPQ) which combines the evaluation of quality preferences and elicitation of idiosyncratic experienced quality factors.

In this paper, we have adopted the ACR-HR (absolute category rating with hidden unimpaired reference video) subjective method which has allowed us to investigate the users' experience of stereoscopic video quality and to compare different coding and transmission scenarios and performances of state of the art video compression standards. Subjective tests in two different laboratories with the ACR method gave quality judgments based on two panels of observers. In addition to answering on a general five point ACR scale, the subjects were asked to indicate visual comfort.

This paper is organized as follows. In Sec. 2, the video preparation processing is described. The subjective experiment is presented in Sec. 3. The results are reported and discussed in in Sec. 4, before concluding the work in Sec. 5.

2. VIDEO PROCESS AND CODING SCENARIOS

2.1 Source videos

In total, 11 source stereoscopic video sequences (SRC) were adopted for the subjective experiment. Each SRC was about 10 seconds long and had higher resolution than SDTV, covering the full range from low motion and low detail through high motion and high detail content. The scenes are summarized in Table 1.

Table 1: Source video sequences

Sequences num.	resolution	frame rate	characterization
src01	1920x1080p	25 Hz	Macro-Recording, time-lapse, surprising motion
src02	1920x1080p	25 Hz	Car racing preparation, high detail, colorful
src03	1920x1080p	25 Hz	Car race, high motion, large depth range
src04	1920x1080p	25 Hz	Animation, human characters, rare colors
src05	1920x1080p	24 Hz	Mesh grid rendering, high detail, small depth range
src06	1920x1080p	24 Hz	Rendered transparent glass ball, circular motion
src07	1280x720p	25 Hz	Group of parachutists, unsteady camera, flapping clothes
src08	1440xs1080p	25 Hz	Market place with groups of people, skin colors
src09	1024x576p	25 Hz	Night scene, fireworks, large depth effects, sudden motion

src10	1024x576p	25 Hz	Uphill hiking group, natural colors, highly detailed trees
src11	1920x1080p	25 Hz	Macro-Recording, time-lapse, large depth perception

2.2 Video processing

Several different scenarios, called Hypothetical Reference Circuits (HRC) according to the terminology of the Video Quality Experts Group (VQEG)[12] were used in creating the Processed Video Sequences (PVS). The SRCs were transformed into PVS according to Figure 1.
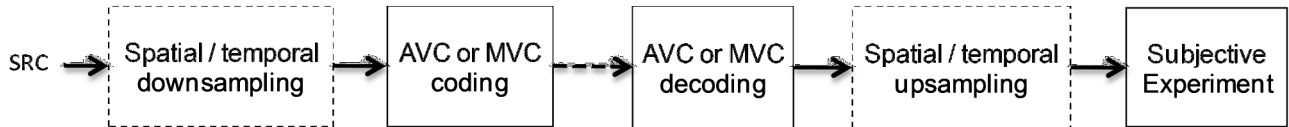

Figure 1: Setup of encoding and processing steps in order to generate the PVS

A SRC was firstly processed with spatial or temporal downsampling (an optional step used for certain HRCs). The spatial downsampling was performed symmetrically using a "lanczos-3 filter" with 1/4 and 1/16 of the original video size; in the temporal downsampling the video frame rate was reduced to 1/2 and 1/3 of the original frame rate, which resulting in approximately 12 and 8 frames per second. Secondly, video encoding process was added to the sequence, the H.264/AVC video encoder in its reference implementation JM 17.0 was used to create the simulcast encoded sequences, and JMVC 7.1 was used to generate multiview encoded videos. After encoding the encoded bitstream was in many cases supposed to be transmitted over a potentially packet loss network, however, it was not the purpose to study such effects in this work, so in our simulation we skipped transmission and directly decoded the bitstreams. The decoded sequences were then 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 we used in the subjective experiment.

Table 2: List of processing conditions (HRC)

HRC Nr.	Codec	QP	Other Degradation
1	none	-	ref. 3D
2	Simulcast	26	/
3	Simulcast	32	/
4	Simulcast	38	/
5	Simulcast	44	/
6	MVC	26	/
7	MVC	32	/
8	MVC	38	/
9	MVC	44	/
10	Simulcast	26	fps/2
11	Simulcast	26	fps/3
12	Simulcast	26	Res/4 (width/2 * height/2)
13	Simulcast	26	Res/16 (width/4 * height/4)
14	Simulcast	-	ref. 2D
15	Simulcast	38	2D

Table 2 lists all HRC conditions. In order to cover the range of typical coding qualities, in HRC1-4 and 5-8 the quantization parameter (QP) was varied from 26 to 44 with a stepsize of six. Incrementing the QP by six, doubles the

quantization stepsize of the linear quantizer for the Discrete Cosine Transform (DCT) coefficients in the H.264 encoder. This also approximately halves the bitrate. Further information can be found in Barkowsky, M.[7]. Please note that the bitrate at the same QP also depends on the properties of the SRC. The fixed QP approach was preferred to fixed bitrate as it helps to cover the full range of quality for each SRC. Temporal and spatial resampling were based on the simulcast coding with QP at 26. HRC 1 is an uncompressed and undistorted video that acted as a reference 3D video to compare to the other conditions. For each source video, a corresponding 2D presentation was also introduced by duplicating the left view video and displaying the same view for the left and the right eye. This provides a pure 2D impression on the 3D screen. All processes were symmetric, which means in for a certain sequences the video processing were equally imposed on both the left and the right view of the side-by-side S3D videos.

3. SUBJECTIVE EXPERIMENT

The subjective experiments were performed at two labs independently: at the University of Nantes IRCCyN, France (Lab 1) and Acreo AB, Sweden (Lab 2). For this cross-lab comparison, the ambient and all hardware and software at both locations were adjusted as similarly as possible. The lab environments adhere to the lab setup defined in the recommendation ITU-R BT.500-11[8]. A 23" Dell LCD display (120Hz, resolution 1920x1080p) was used for displaying 3D videos in the experiment together with a pair of active shutter glasses from the Nvidia 3D vision system. The display was positioned far enough from the wall to avoid any conflicts of the displayed 3D content with the real world. The viewing distance was set to 3 times of the display height which is the same value used in the VQEG HDTV testplan[12]. The maximum crossed and uncrossed disparity of each SRC sequence was manually determined in order to assure that the videos were displayed in the comfortable viewing zone described in Chen, W.(2010)[13]. The voting interface for the viewers to rate the video quality was shown on a separate display. At IRCCyN, backlighting was used for adjusting the lab luminance, the luminance level of the reflection from the gray wall behind the screen was set to 50cd/m^2 which corresponded to 15% of the peak luminance of the display after passing the shutter glasses when they were activated. At Acreo the room illuminance when displays were on was set to 20 (lux) which was close to darkness for the sake of avoiding reflections from objects other than the display. No flickering was perceived in any of the laboratories. The video sequences were displayed in uncompressed format in order to make sure that all observers were given the same presentation of the same video sequence. In order to assure that no temporal distortion was introduced by the player, the videos were preloaded into the computer's Random Access Memory (RAM) and special care was taken that the playout of twice the Full-HD resolution was performed without temporal jitter.

Figure 2: Subjective experiment rating interface

A training session containing 10 sequences was pre-conducted before the formal rating session so that observers would become accustomed to the PVSs' characteristics and the rating interfaces. Both training session and rating session were using the "absolute category rating with hidden reference" (ACR-HR) method in our experiment; hence the observers had a smooth transition from training session to rating session without feeling any boundaries. The PVSs were presented in random order and they were rated independently on the ACR category scale which is five-point quality scale defined

by ITU[9] as it shown in Figure 2 (Excellent, Good, Fair, Poor and Bad, which are later mapped to the scores 5, 4, 3, 2, and 1 respectively). The subjective test instructions, questionnaires as well as rating interface were presented in the observers' native language (French in the University of Nantes, and Swedish in Acreo, see table 3), for other international observers English were used.

Table 3 displays a French and Swedish version of the subjective experiment user interface which corresponds to the English version shown in Figure 2.

Table 3: French and Swedish version of subjective experiment user interface

		French	Swedish
Title		Comment jugez-vous la qualité d'expérience dégagée par la séquence 3D?	Vad tycker du om kvaliteten i denna 3D presentation?
part1		Qualité d'expérience	Upplevelsekvalitet
		[Excellente, Bonne, Assez Bonne, Médiocre, Mauvaise]	[Utmärkt, Bra,Varken eller, Dålig,Usel]
part2		Confort visuel	Visuella bekvämligheten
		[Beaucoup plus…, plus…., Aussi…., Moins…, Beaucoup moins….]	[Mycket mera…, Mera…, Lika…, Mindre…, Mycket mindre…]
		….confortable à regarder que la télé 2D	…..visuellt bekvämt än att se på 2D TV
Validate		Valider	Validera

For each sequences, besides the evaluation for the overall video quality of 3D experience, we included a visual comfort comparison scale to evaluate the visual comfort associated with the visualization of the sequences compared to viewing on a conventional 2D television. The next presentation followed immediately[†] after an observer validated his / her vote on both quality of experience and visual comfort. The average time for rating a single PVS and preloading the next sequence was about 16 seconds, thus about one third of the time needed for the same experiment if it were proceed with DSCQS.

The subjective experiment contained a total of 175 videos: 10 training sequences, 11 references and 154 impaired sequences which were presented in pseudo-random order. Prior to the subjective experiment, the observers were screened for visual acuity using a Snellen Chart, stereoscopic acuity using a Randot Stereo test and color blindness. The whole experiment was divided into two sessions of approximately 50 minutes each with pauses after about 15 minutes of viewing time. In total 48 naïve observers (24 at each lab) participated the subjective experiment. After the experiment all observers' votes were screened according to ITU-R BT.500 and the VQEG HDTV testplan, 7 (2 from IRCCyN and 5 from Acreo) were rejected. The remained 41 observers consisted of 21 male and 20 female with an average age at 27.9 years old (minimum 15, median 23, maximum 64).

4. RESULTS

4.1 Cross lab Comparison

Figure 3 is a scatter plot of the subjective experiment data collected from two labs. All PVSs were mapped in horizontal axis with Lab 1's MOS scale and in vertical axis with Lab 2's voting scale. It shows the MOS results from the two laboratories have a similar trend, though the experiments were done in different locations and different observer groups. The green diagonal line is a reference. It indicates the ideal case in which the data from the two laboratories would match perfectly to each other. However, the real data had a small deviation downside of the diagonal line, which mean there was a difference between the data of two labs, and the observers in Lab1 (IRCCyN) were giving higher score than in Lab 2 (Acreo) for the same PVS in most of the cases. In fact the MOS from Lab 1 had slightly larger span from lowest 1.7 to highest 4.4 comparing to Lab 2 which spans from 1.9 to 4. The blue regression line is Lab 1's data mapped to Lab 2's, and the black regression line shows lab 2's data mapped to lab 1 which was actually used in the final combined data analysis.

An ANalysis Of VAriance between groups (ANOVA)[14] was performed with the laboratories as one between factor and 11 SRCs times 15 HRCs as within factors. This shows that there was a significant difference in the main effect of the laboratories $F_{(1, 41)} = 5.83$ (Fisher-Snedecor distribution), the significance level or critical p-value $= 0.02 < 0.05$, corresponding to a 2% chance of rejecting the null hypothesis when it is true. The null hypothesis here means the distribution from the two labs stems from the same statistical process.

[†] There were sometimes a small delay due to the video loading time
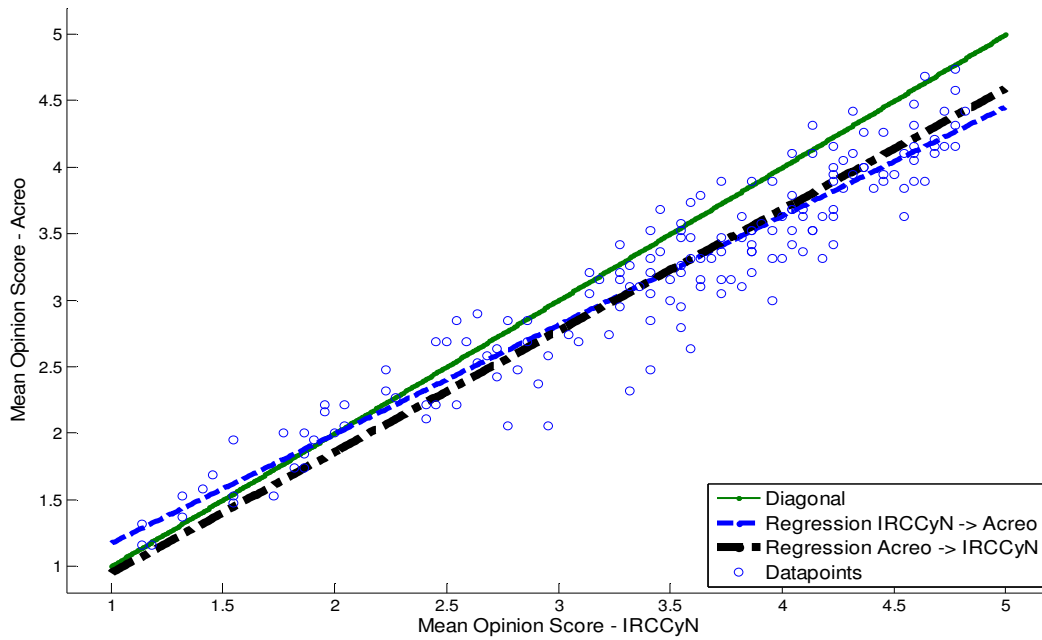
Figure 3; Scatter plot of data from two labs with linear regression

However if we applied the linear transformation of Lab 1's to Lab 2's data by formula 1, most of the difference between the laboratories can be taken out. In other words, we kept lab 1's data untouched and rescaled lab 2's data to match lab 1. The choice was completely arbitrary and could equally well have been done the other way around.

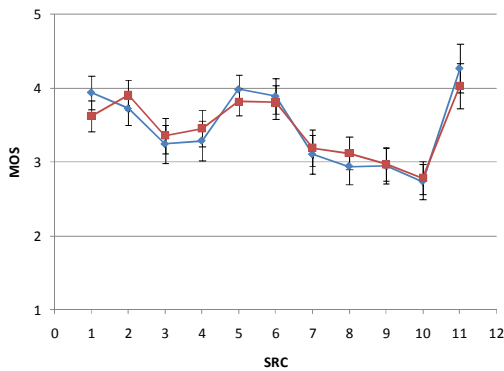$$Acreo\ realigned\ data = 1.22 \times Acreo\ original\ data - 0.43 \tag{1}$$

Figure 4: The MOS across SRC of the different laboratories after scaling Lab 2's (Acreo) data to Lab 1 (IRCCyN). The error bars shows 95% confidence intervals
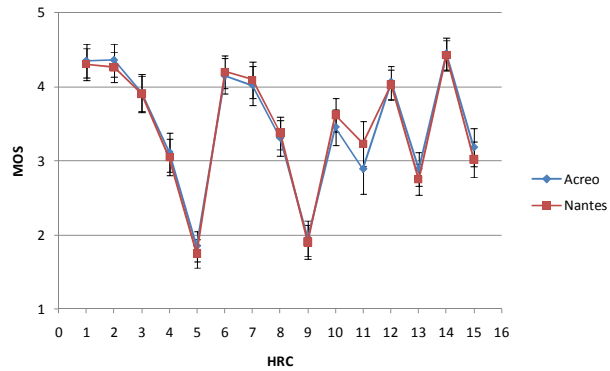

Figure 5: The MOS across HRC of the different laboratories after scaling Lab 2's (Acreo) data to Lab 1 (Nantes). The error bars shows 95% confidence intervals

We then analyzed the data again with the above mentioned ANOVA. This time the main effect of laboratories was not significant $F_{(1, 39)} = 0$, $p = 0.99 > 0.05$. The main effects of SRC and HRC were significant with $F_{(10, 390)} = 68.80$, $p = 0.00 < 0.05$ and $F_{(14, 546)} = 197.16$, $p = 0.00 < 0.05$ respectively. The interaction between SRC and HRC was also significant $F_{(140, 5460)} = 6.70$, $p = 0.00 < 0.05$. The interactions of SRC with laboratories, see Figure 4, was also significant $F_{(10, 390)} = 2.52$, $p = 0.01 < 0.05$. The interaction between the HRC and laboratories, see Figure 5, was however not significant. Therefore the two tests were combined and analyzed as a single evaluation after aligning the data from one lab to the other. Such an alignment is often necessary between two laboratories performing the same experiment, in particular if the language is different. The notion of the absolute categories has slight offsets in different languages.

4.2 Analysis of observers' experience of 3D video quality

The average MOS as shown in Figure 4 were calculated across all 11 SRCs. The error bars indicate 95% confidence intervals. It can be seen that the MOS varies for different video source content. In general, 3D videos with the original Full HD resolution (SRC1~6 and 11) were rated better than the others (SRC7~10). The videos with lower original resolution were upscaled to Full HD resolution for displaying on the screen, hence it is reasonable to expect a lower MOS. Apart from the original resolution effect, SRC 10 was rated lowest, which may be because this 3D content had the highest crossed and uncrossed disparity among all SRCs. The content was highly detailed and the video was shot in bright sunlight, thus the background of the scene appears very bright as well. On the other end of the scale, SRC11 was rated the topmost of the visual quality. This content had a good depth presentation, slow motion and the background is dark.

Figure 5 presents the MOS partitioned by the 15 HRCs. Obviously the degree of compression during the video encoding expressed as the Quantization Parameter (QP) has great impact on the MOS. Both H.264/AVC simulcast and MVC show a similar trend, the assessed quality level decreases when the encoding QP was increased. HRC12 (spatial resolution divided by 4) got a better MOS (4, "good") than the other 3 temporal and spatial processes. It shows the statistically the same MOS as the 3D encoding only scenario with QP32.

This is particularly true for those SRC that have Full-HD resolution. A detailed analysis per SRC showed that people did not notice the resolution reduction by a factor of 4. For the Full-HD SRCs, the MOS value of the reference videos was equal to 4.4, while the MOS value for the resolution reduction case was equal to 4.3. This was tested to be statistically insignificant. MOS of the temporal processes dropped moderately (3.5 for FP/2, and 3.1 for FP/3), which is about the similar quality level as coding only case at QP38. Resolution/16 reduction got the worst quality evaluation with MOS= 2.8.

As we have seen from our previous work [15], in an ACR subjective experiment, the 2D unimpaired reference presentation (HRC14 in Figure 5) may get slightly better MOS than the 3D unimpaired reference sequence (HRC1) even if this absolute difference is small. By looking further at the visual discomfort data collected during the subjective test, the discomfort values of HRC1 and HRC14 rated in a MOS scale were statistically equal, with a vote of approximately 3.5, indicating a value that may be translated to "slightly more comfortable than 2DTV". Hence the visual discomfort may not have been the factor which affected the 2D preference. However this characteristic was largely dependent on the 3D video source, e.g. for SRC11, the 3D reference videos were clearly rated superior to the 2D case. We also noticed that observers' preference of 2D video disappeared when the 2D video was heavily compressed (HRC15), in contrast with HRC4-AVC encodedat QP38, they both have a similar MOS=3.

4.3 Bitrate / MOS analysis

Figure 6 plots the average MOS versus bitrate in a semi-logarithmic scale. All SRCs are compared individually for the coding performance of H.264 simulcast and MVC. The blue and green dashed lines connect the 4 data points from the experimental data of MVC and H.264 simulcast coded videos respectively. These points present QP 44, QP 38, QP 32, QP26 that were related to their corresponding bitrates along the X axis. The figure shows that the MVC and H.264/AVC curves are quite close, however the MVC performed slightly better than H.264 simulcast. It can also be noticed that the gain decreases with higher bitrates. For most of the SRCs the QP32 and QP26 are statistically indistinguishable as the curves show a flat out trend at the top.
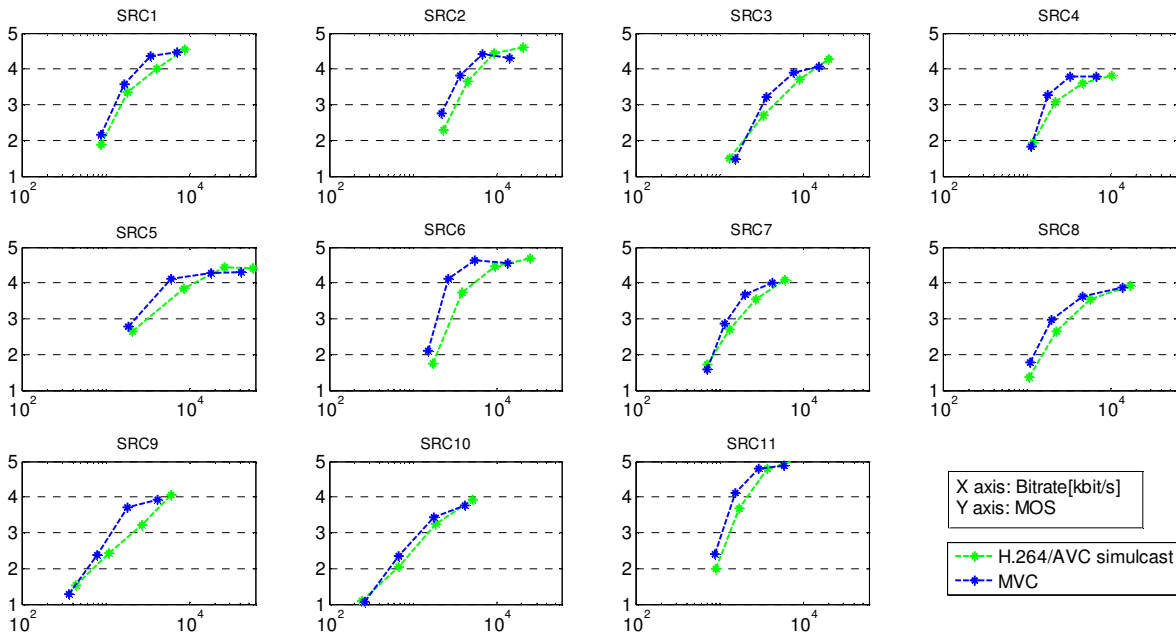
Figure 6: Bitrates versus MOS of MVC and H.264 simulcast comparison

In order to have a simple way of comparing between the cases of temporal and spatial resolution reduction to the coding only cases we introduce two factors: "bitrate gain" and "quality gain". They are demonstrated by an example in Figure 7 and equations 2 and 3.
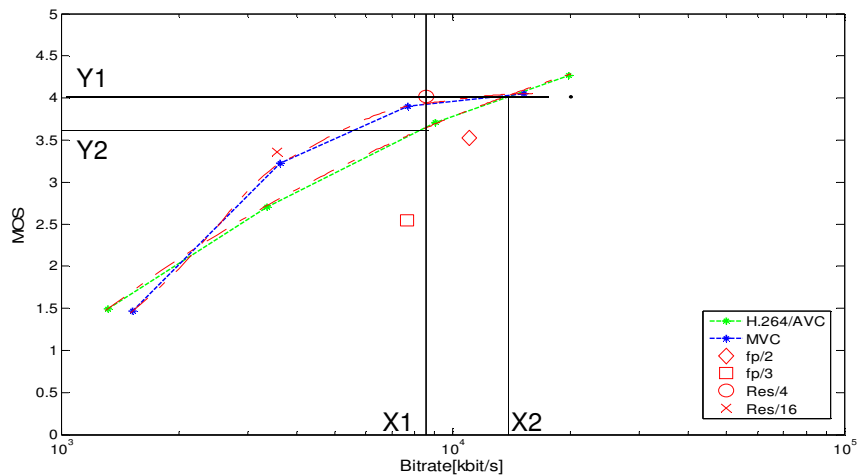

Figure 7: An example of comparison of different video processing in terms of bitrate efficiency

Figure 7 plots the average rating of SRC3 for H.264 simulcast coding, MVC coding, temporal and spatial artifacts as a function of bitrate. The MVC and H.264 simulcast lines were further interpolated and extrapolated by the red curve with curve fitting tools. The HRCs of frame rate and resolution reduction are indicated by 4 individual data spots. Figure 7 shows an example calculation of the two indicators for the "resolution /4" reduction to H.264/AVC coding only cases. First, we obtained the "resolution /4" data point (x1,y1), and we drew a vertical line and a horizontal line. These lines intersect with the rate distortion curves of the coding only cases. The vertical line generates an intersection point with the H.264/AVC interpolated line at (x1, y2), and the horizontal line crosses the H.264/AVC line at (x2, y1). In some cases, "resolution/4" has a higher MOS than the MVC or the H.264/AVC curves and the horizontal line passing through this point didn't intersect with the MVC or the H.264/AVC curves. This was particularly the case when they showed a flat out trend at higher bitrates. In these cases we used the bitrate of QP26 encoded points on the MVC or simulcast curves instead in order to obtain the coordinate x2. The "bitrate gain" factor is given by formula 2, which indicates the amount

of bitrate that can be saved when the MOS remains constant, i.e. the service provider offers a guaranteed quality of 3DTV services.

$$Bitrate\ Gain = \frac{X1}{X2} \qquad (2)$$

The "quality gain" factor is defined in equation 3. It indicates for a given bitrate limit the quality gain that can be achieved by a resolution reduction of 4. This is a scenario in which the 3DTV service provider offers a fixed access bandwidth to the subscriber.

$$Quality\ Gain = Y1 - Y2 \qquad (3)$$

By applying these two factors to all temporal and spatial processed points for each SRCs we can get Table 4 for the bitrate gain comparison and Table 5 for quality gain comparison.

Table 4: Comparison of spatial and temporal reduction performance to 3D coding concerning bitrate gain at same quality level

bitrate gain	FP/2	FP/3	Res/4	Res/16	FP/2	FP/3	Res/4	Res/16
	comparing with H.264 simulcast				comparing with MVC			
src1	122,5%	156,5%	59,8%	68,7%	221,8%	226,1%	110,5%	92,1%
src2	379,1%	302,4%	75,4%	77,4%	502,5%	400,7%	57,9%	105,5%
src3	148,1%	264,8%	63,1%	57,5%	231,5%	329,4%	72,3%	89,0%
src4	236,3%	300,2%	39,1%	78,3%	376,3%	388,4%	59,7%	104,5%
src5	348,6%	324,3%	58,8%	66,9%	822,0%	729,7%	32,5%	140,3%
src6	537,1%	536,1%	30,7%	78,0%	899,5%	784,7%	56,3%	127,9%
src7	169,2%	240,0%	63,4%	133,5%	248,9%	291,4%	96,1%	134,4%
src8	447,9%	490,1%	76,5%	106,6%	641,8%	648,9%	120,0%	137,4%
src9	118,2%	92,0%	73,4%	158,7%	261,1%	195,7%	159,3%	146,4%
src10	165,4%	289,6%	136,5%	183,5%	230,6%	376,6%	177,0%	172,7%
src11	194,3%	182,7%	61,4%	69,4%	281,0%	272,8%	35,0%	102,5%
mean	260,6%	289,0%	67,1%	98,0%	428,8%	422,2%	88,8%	123,0%

Table 4 gives an overview of bitrate gain of all frame rate and resolution reductions. The cells marked with green shade indicate when the corresponding process is efficient and thus saves bandwidth compared to MVC or H.264 simulcast coding. For the other cells marked with red where MVC or H.264/AVC performs better without the preprocessing of spatial or temporal downsampling. Please note, that the resolution reduction cases were only performed with H.264 simulcast coding so the comparison to MVC coding may be misleading and higher gains may be achieved when the downsampled videos are later encoded with MVC instead of H.264/AVC.

Comparing the bitrate gain of spatial and temporal resolution reduction to the H.264/AVC encoding, the "resolution / 4" is clearly superior to others with a mean value of 67%, i.e. on average, "resolution / 4" process only uses 67% of the bandwidth which was needed for transmitting H.264 simulcast coded 3D videos at the same visual quality level. There is an exception for SRC10 where all spatial and temporal processes cost more bandwidth than the coding only cases. The "resolution/16" reduction only works efficient among the videos with originally full HD resolution (using only 60% of bandwidth in average). For the videos with a lower resolution, the transmission at the original resolution with H.264 is more efficient. However, when the comparison is extended to MVC, we see that "resolution/4" outperforms MVC among most of the full HD content videos. Previously we have already seen that MVC performs slightly better than H.264 in general. Table 4 also reveals the temporal reduction didn't save any bitrates at all, on the contrary it required 2.6 (for 12fps) and 4 times (8fps) more bandwidth compared to H.264 simulcast and MVC.

Table 5 shows a comparison of the quality gain based on the same service bandwidth. The results are similar results to those of the bitrate gain. The "resolution/4" wins in the quality gain with an average improvement of 0.24 MOS compared to the H.264/AVC encoding only case. For the videos with higher resolution the "resolution/16" can get a quality gain of 0.53 MOS. In the comparison to the MVC case, only the "resolution/4" shows a few small enhancements (0.10 MOS) for the higher resolution contents.

Table 5: Comparison of spatial and temporal reduction performance to 3D coding concerning quality gain with constant bitrate limit

quality gain	FP/2	FP/3	Res/4	Res/16	FP/2	FP/3	Res/4	Res/16
	comparing with H.264 simulcast				comparing with MVC			
src1	-0,17	-0,33	0,40	0,55	-0,36	-0,70	-0,02	0,16
src2	-1,33	-1,43	0,12	0,52	-1,11	-1,38	0,07	-0,10
src3	-0,33	-1,00	0,37	0,57	-0,47	-1,29	0,12	0,13
src4	-0,37	-0,80	0,35	0,39	-0,45	-0,92	0,12	-0,10
src5	-0,57	-0,78	0,29	0,38	-0,43	-0,69	0,15	-0,45
src6	-1,12	-1,64	0,32	0,53	-1,10	-1,67	0,09	-0,60
src7	-0,44	-0,96	0,37	-0,48	-0,64	-1,26	0,02	-0,79
src8	-0,98	-1,30	0,16	-0,10	-1,00	-1,37	-0,08	-0,57
src9	-0,17	0,08	0,30	-0,41	-0,54	-0,54	-0,43	-0,55
src10	-0,40	-1,00	-0,33	-0,57	-0,38	-1,03	-0,56	-0,76
src11	-0,52	-0,81	0,24	0,80	-0,49	-0,85	0,15	-0,05
mean	-0,58	-0,91	0,24	0,20	-0,63	-1,06	-0,03	-0,33

As a conclusion, it may be stated that for an HD 3DTV transmission system, actually a reduction of the resolution by a factor of four before the video encoding will result in a better quality. It will not only help the service provider to save the bandwidth but also to save some amount of hardware processing which would be needed for encoding and decoding of two times a Full HD video.

4.4 Gender analysis

An Anova was performed with gender as one between factor and 11 SRCs times 15 HRCs as within factors. The proportion of women to men is about one to one in the subjective test (21 male and 20 female). This analysis shows that there was no main effect difference between the genders. However, the interaction between gender and HRC was significant, see Figure 8, $F_{(14, 546)} = 2.38$, $p = 0.00 < 0.05$. The interaction with the SRC was not significant, see Figure 9. We can find from Figure 8 that female observers were voting slightly more positive, and the MOS spans a bit larger range compared to the male observers.
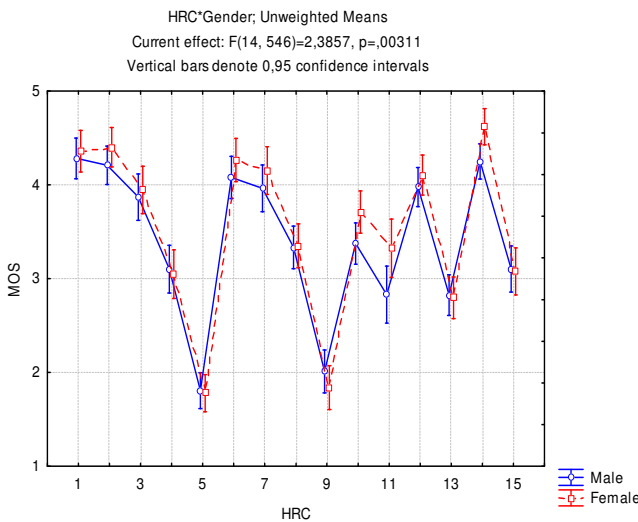

Figure 8: The mean MOS across HRC for the different genders. The error bars shows 95% confidence intervals
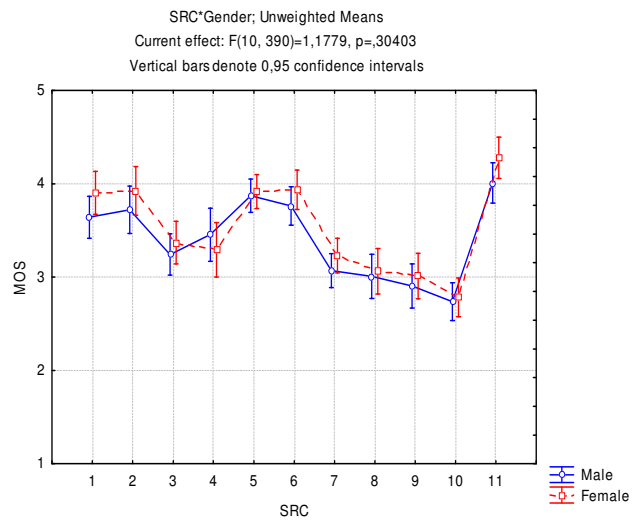

Figure 9: The mean MOS across SRC for the different genders. The error bars shows 95% confidence intervals

5. CONCLUSIONS AND FUTURE WORK

In our study we evaluated several video processing methods as well as the state of the art 3D coding standards. The results were based on a subjective experiment using ACR of S3D videos performed in two laboratories, one in France and one in Sweden. There was a difference between the laboratories that could be modeled well by a linear transformation. Even after transformation of one of the laboratories data with the linear transformation, a significant interaction between the laboratories and the SRCs remained. It was also discovered that there was a significant interaction between the genders and the HRC i.e. females voted more positively on high quality videos than males, but also decreased their quality ratings faster that males.

We presented a specific focus on bandwidth saving using the subjective experiment results. We discovered that a resolution reduction of four may result in higher bitrate efficiency when H.264 video coding is used. The result that this scenario is also preferable to MVC coding suggests that the reduction of the spatial resolution may be advantageous for MVC as well. It was seen that MVC performs better than H.264 simulcast, however this advantage decreases when the bitrate is higher. The reduction of the frame rate did not save bandwidth but it actually reduced the video quality to a large amount. It was noted that the 3D video content is an important and largely independent factor; it has strong effects on the video quality, the visual comfort and the bitrate. The results can be applied to an HD 3DTV transmission system, as the resolution reduction not only helps service provider to save bandwidth but also to save some amount of hardware processing which would be needed for encoding and decoding.

6. ACKNOWLEGEMENT

REFERENCES

[1] ITU-T Rec. & ISO/IEC 14496-10 AVC, "Advanced video coding for generic audiovisual services", ITU-T P. (2005).

[2] Vetro,A., Yea,S., Chen,Y., Shimizu,S. Pandit,P, Lim,C.; "Status of MVC text, software, and conformance" JVT-AE014, (2009).

[3] Saygili, G., Gurler, C.G., Tekalp, A.M., "3D display dependent quality evaluation and rate allocation using scalable video coding", Image Processing (ICIP)IEEE, page(s)717 - 720, (2009)

[4] Ozbek, N.; Tekalp, A.M.; Tunali, E.T.; , "Rate Allocation Between Views in Scalable Stereo Video Coding using an Objective Stereo Video Quality Measure," IEEE International Conference on Acoustics, Speech and Signal Processing, (2007).

[5] Stelmach, L.; Wa James Tam; Meegan, D.; Vincent, A.; , "Stereo image quality: effects of mixed spatio-temporal resolution", IEEE Transactions on Circuits and Systems for Video Technology, vol.10, no.2, pp.188-193, (2000)

[6] Pinson, M., Wolf, S. and Cermak,G., "HDTV subjective quality of H.264 vs. MPEG-2, With and Without Packet Loss," IEEE Transactions on Broadcasting (2010).

[7] Barkowsky, M., M. Pinson, R. Pépion and P. Le Callet, "Analysis of Freely Available Dataset for HDTV including Coding and Transmission Distortions". Fifth International Workshop on Video Processing and Quality Metrics (VPQM), (2010).

[8] ITU, "Methodology for the subjective assessment of the quality of television pictures," in Recommendation BT 500-11, ed: International Telecommunication Union., 2002.B

[9] ITU-T Study Group 12. "ITU-T P.910 Subjective video quality assessment methods for multimedia applications". ITU-T P.910, (1997).

[10] Merkle, P.; Smolic, A.; Muller, K.; Wiegand, T.; , "Efficient Prediction Structures for Multiview Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, vol.17, no.11, pp.1461-1473, Nov. (2007).

[11] Strohmeier,D. "Open Profiling of Quality as a mixed method approach to study multimodal experienced quality", 8th European Conference on Interactive TV and Video EuroITV, workshop "Methods for User Studies of Interactive (TV) Technologies" (USIT), (2010).

[12] 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).

[13] Chen, W., Fournier, J. Barkowsky, M. and Le Callet, P. "New Requirements of Subjective Video Quality Assessment Methodologies for 3DTV", Fifth International Workshop on Video Processing and Quality Metrics (VPQM), (2010).

[14] Spiegel, M. R., & Stephens, L. J.; "Schaum's outline of theory and problems of statistics". McGraw Hill, (1998).

[15] Barkowsky,M., Wang, K., R Cousseau,R., Brunnström, K. Olsson,R., Le Callet,P. "Subjective Quality Assessment of error concealment strategies for 3DTV in the presence of asymmetric transmission error", IEEE conference Packet Video Workshop in HongKong, (2010).