



HAL
open science

INFLUENCE OF DEPTH RENDERING ON THE QUALITY OF EXPERIENCE FOR AN AUTOSTEREOSCOPIC DISPLAY

Marcus Barkowsky, Romain Cousseau, Patrick Le Callet

► **To cite this version:**

Marcus Barkowsky, Romain Cousseau, Patrick Le Callet. INFLUENCE OF DEPTH RENDERING ON THE QUALITY OF EXPERIENCE FOR AN AUTOSTEREOSCOPIC DISPLAY. International Workshop on Quality of Multimedia Experience, Jul 2009, San Diego, California, United States. pp.6. hal-00433915

HAL Id: hal-00433915

<https://hal.science/hal-00433915>

Submitted on 20 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INFLUENCE OF DEPTH RENDERING ON THE QUALITY OF EXPERIENCE FOR AN AUTOSTEREOSCOPIC DISPLAY

Marcus Barkowsky, Romain Cousseau, Patrick Le Callet

IRCCyN UMR 6597 CNRS, École Polytechnique de l'Université de Nantes,
rue Christian Pauc, La Chantrerie 44306 Nantes, France
{Marcus.Barkowsky,Romain.Cousseau,Patrick.LeCallet}@univ-nantes.fr

ABSTRACT

Autostereoscopic displays simplify the presentation of 3D content because they do not require any glasses and they allow the perception of motion parallax. While the perception of depth is certainly an added value, the technical rendering process of the current display technology also introduces artifacts. For the viewer, the trade-off may be expressed in terms of quality of experience. However, quality of experience assessment related to 3D is still an open issue. Towards this goal, several original subjective test methods are proposed and compared that are meant for assessing the quality of experience. A split-screen setup simultaneously displays a 2D and a 3D presentation. The observers vote according to their preference in terms of quality of experience. In four experiments, the influence of the depth rendering process is evaluated. The results indicate that the degradation by the depth rendering process may easily dominate the added value of depth in a content specific manner.

Index Terms— Video signal processing, autostereoscopic display, quality of experience, subjective testing, 3D video quality

1. INTRODUCTION

In recent years, considerable effort has been made to propose 3D television as one of the next milestones in broadcast applications. The backwards compatible broadcast using anaglyph stereoscopic television has already been tried in the last century but it was not accepted by the customer. Due to the new transmission systems for broadcasting, the upcoming system does not need to be backwards compatible and the whole process chain can be reconsidered. The analysis starts at the production of content, includes the transmission channels and ends at the display and the viewer.

The representation of the content during the transmission and processing stages may be changed as well. Consequently, the anaglyph stereoscopic transmission has been widened to multiview transmissions. Multiple views can be used to support additional depth cues, especially the motion parallax effect [1].

For displaying the multiview content, an autostereoscopic multiview display is very well suited. It does not require the user to wear anaglyph, polarized, or shutter glasses and the motion parallax effect can be exercised by several observers simultaneously. Currently, most autostereoscopic displays are lenticular displays which use a tilted lens array in front of a TFT screen to project the different views. This leads to a significant reduction in image resolution and the displays also suffer from a considerable amount of crosstalk between the individual views.

One of the most prominent displays in this category is the Philips 42-inch 3D display which is used for the experiments in this paper.

It features nine different views that are internally generated from the 2D texture and additional depth information at its input. This display has already been used in several subjective experiments, for example, in [2] the effect of depth map compression using H.264 coding was subjectively tested.

The measurement of subjective quality on a conventional 2D display has been standardized for many different scenarios, e.g. standard definition television [3] or multimedia [4]. For three dimensional presentations, the assessment is more difficult because most observers are only used to two dimensional presentations. The experience of having different views for each eye and the perception of depth is exciting for the viewer at first but the effect on the individual quality of experience is usually not normalized. When the subjects are asked to rate the video quality in a single stimulus test on an absolute scale they often only rate the 2D video quality. In [5], the concept of “naturalness” was introduced based on a subjective still image test with the Philips autostereoscopic display. When the participants were asked to rate the absolute quality on an ACR scale based on this attribute, they considered the added value of depth to some extent.

In this paper, a different approach is tested: The observers compare two presentations side by side and they decide whether they prefer left or right in terms of quality of experience. Preliminary experiments demonstrated that the displayed 3D image appeared to be blurred compared to the texture information at its input, i.e., relative to the 2D presentation. Thus, even if there was no additional distortion in the texture or in the depth map, the 3D presentation is affected. The degradation is correlated with the values of the depth map at the corresponding spatial location. When the depth plane corresponded to the display plane, such that the content did not appear to be either in front or behind the display, the effect was minimized.

The influence of the depth induced blur effect on the quality of experience is analyzed in a subjective test. Several different protocols are used and compared in order to learn about their suitability for the measurement of quality of experience as opposed to quality of 2D video presentation.

The characterization of the depth induced blur will be performed by comparison with known degradations. This comparison approach is universally applicable to any 3D display solution. Specifically, it can be applied if there is a hidden rendering step included in the process, e.g. in order to convert 2D plus depth to multiple views.

The paper is organized as follows. In Sec. 2 a systematic view of the transmission and display chain will be presented. The subjective test environment and the displayed content are described in Sec. 3. The results of the different evaluations are provided in Sec. 4. Finally, the conclusions are drawn in Sec. 5.

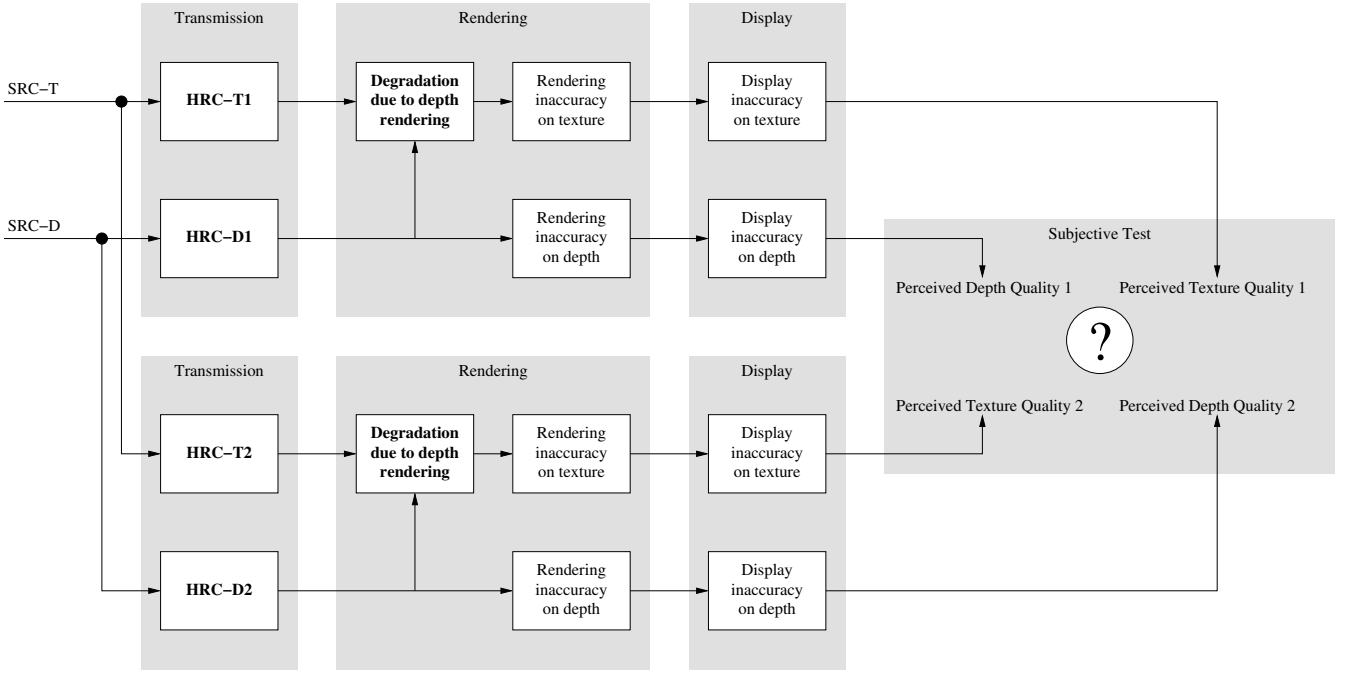


Fig. 1. Transmission and display system overview

2. ANALYSIS SYSTEM OVERVIEW

The video quality that is perceived by the observer may be degraded at different stages in the transmission and display chain. This will be further explained based on Fig. 1.

In our scenario, the 3D source reference channel or circuit (SRC) splits into the texture part which contains the well-known 2D video and the depth information which is represented in a gray level video. Each part has an image resolution of 960x540 pixels at 8 bits per channel.

During the transmission, different processing steps may be applied to the texture part and to the depth information part. This fact is indicated in the block diagram by two different hypothetical reference circuit (HRC) which are termed HRC-T1 and HRC-D1 for the texture and the depth part respectively.

As mentioned earlier, a strong influence of the depth information on the texture was observed which may be assigned to the rendering process. This is modeled in the block *degradation due to depth rendering*. Afterwards, the texture and depth information is prepared for the display. This includes a rendering of nine different views. The resolution of the nine views is lower than the input information, the number of pixels for each view is approximately half of the texture resolution. Additional degradations result from the disocclusion of objects in the views far from the center view that become visible from a certain viewing angle but no texture information may be available to fill the disoccluded area.

Instead of representing the nine views as individual signal paths in the block diagram, it was decided to keep the texture and depth sensation equivalent. This simplification facilitates the analysis of the final perception. The impact of the rendering process on the video quality is included in the steps *rendering inaccuracy of texture and depth*.

The lenticular display is the final step in the processing chain. In

addition to having all deficiencies known in conventional displays, it exhibits crosstalk between different views and color aberration because of the lenses. The influence of those artifacts are modeled in the steps called *display inaccuracy on texture* and *display inaccuracy on depth*.

The resulting sequence is judged by the viewers in a subjective test. A pilot study indicated that a comparison setup as shown in the figure is preferable to an absolute rating scale because the simultaneous presentation allows the observers to concentrate on their overall preference in terms of quality of experience.

3. SUBJECTIVE TEST ENVIRONMENT AND SOURCE MATERIAL

The subjective test was performed in a room conforming to ITU-R BT.500 [3]. The display was connected to a ClearView Server using a Gefen SDI to DVI converter which also doubled the frame rate from 30 fps to 60 fps. The viewing distance was 3m which was specified as the optimal viewing distance for this multiview display. A split screen setup with no mirroring was used for the subjective test because only one display was available. Each sequence pair occurred twice during the subjective test in order to reverse the sides.

A short description of the SRCs is shown in Table 1. Each sequence contained 100 frames. In a pilot study, it was noticed that for the split-screen setup, a short sequence which was played in loop-mode allowed the most stable results. Half of the sequences contained natural content and the other half computer generated content. The quality of the 2D part of the reference sequences was excellent, even though some of the sequences were previously encoded. The depth quality depends on the generation: It was excellent for the computer generated content but contained artifacts from the estimation process for the natural content in SRC4 to SRC6.

Table 1. Reference Sequences

| SRC | | Properties |
|-----------------|---|---|
| Synthetic video | 1 | Big Buck Bunny [6], jump roping, animated movie Depth map from animation |
| | 2 | Logo, rotating items in back, text in foreground Depth map from animation |
| | 3 | Quake3, a fast run with two opponents Depth map rendered via OpenGL control |
| Natural video | 4 | Watermill, man standing at a workbench Depth map generated from 2D sequences |
| | 5 | Breakdancers, fast motion, 15fps Depth map generated from multiview sequences |
| | 6 | Mont Saint Michel, street view with people moving Depth map generated from stereo images |

4. TEST DESCRIPTION AND RESULTS

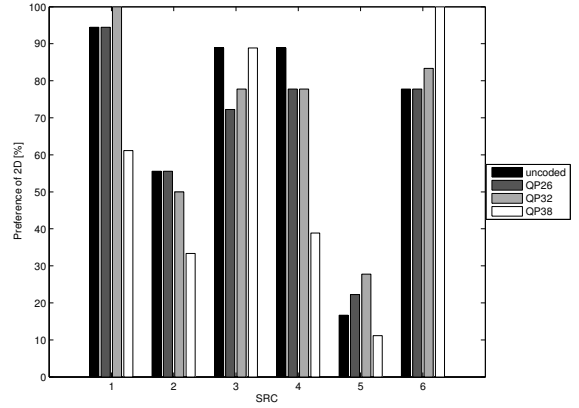
As already mentioned in Sec. 1, the depth rendering causes an additional distortion to the texture on the screen. In particular, if the depth map indicates that some part of the content shall be displayed in front of the screen or behind the screen, a blur is perceived for that part of the content. The best rendering quality is achieved in the central plane, and with a larger difference from the central plane the display exhibits a larger blur effect.

By setting the depth map sequence to a constant value of 128, the display is forced to render the texture in the central plane. In this case, no 3D effect is visible but the texture quality is highest. This setup with a constant depth value will be referred to as 2D presentation in this paper. It virtually removes the step called *degradation due to depth rendering* in Fig. 1.

It shall be noted that the quality is still lower than on a comparable 2D display because the lenticular array was always activated. The display features a special 2D mode which reduces the influence of the lenticular display technology but this mode was not used because some parts of the screen always displayed 3D content.

Three different types of test setups were used: “Pair comparison”, “search for equal quality of experience” and “search for the best quality of experience”. As will be explained in detail later, the subjective test using the “Pair Comparison” protocol always presented the same sequence in 3D presentation on one side of the display and in 2D presentation on the other side. The two conditions could be easily distinguished by the viewer. If these presentations had been done in a dedicated subjective test, the viewer might have decided to always prefer 2D or 3D presentation. By combining the different subjective tests, this effect was avoided. A dedicated graphical user interface application was used by the observers on an additional LCD monitor. All presentations from all subjective tests were randomized for each observer implying the constraint that the SRC and the HRC was different for two subsequent votes. The drawback of combining the different test protocols is that the voting interface changed with each video presentation.

Because this setup needs a high level of concentration throughout the subjective test, we decided to perform an expert-only subjective test and we confirmed that the participants understood the necessity to pay close attention to the type of question asked. Nine experts on video coding and quality assessment participated in the subjective test. All observers passed the test for visual acuity and the randot stereo test. Note that due to the side-by-side setup and the reversal of the sides, each expert voted twice on each condition.

**Fig. 2.** Bar chart for experiment 1

4.1. Experiment 1: Preference of 2D versus 3D

The first experiment displayed 2D on one half of the screen and 3D video on the other half. Thus, HRC-D1 contained the constant depth 2D display HRC and HRC-D2 did not exist. In the first condition, the texture was not processed, thus HRC-T1 and HRC-T2 did not exist.

Additionally, the effect of coding degradations on the preference of 2D versus 3D was tested. In three additional conditions, the 2D part of the sequence was encoded using the ITU-T H.264 [7] reference software in the version JM 15.1. The frame type was chosen as IBBPBB... and an immediate decoder refresh (IDR) picture was forced after 24 frames. The quantization parameter (QP) was fixed for all slice types which leads to a variable bitrate but the quality is comparable for different content. Three different QP values were tested: 26, 32 and 38 which corresponds approximately to “very good”, “good” and “average” quality in a subjective test on IPTV material when transmission distortions are included.

Thus, the same texture information was displayed on both parts of the screen but the depth information was changed. The subjective test protocol was “Pair Comparison” as described in ITU-T P.910 [4].

Table 2. Results for experiment 1: Preference of 2D on 3D display in percent

| SRC | uncoded | QP26 | QP32 | QP38 | Avg. |
|------|---------|------|------|------|------|
| 1 | 94 | 94 | 100 | 61 | 87 |
| 2 | 56 | 56 | 50 | 33 | 49 |
| 3 | 89 | 72 | 78 | 89 | 82 |
| 4 | 89 | 78 | 78 | 39 | 71 |
| 5 | 17 | 22 | 28 | 11 | 20 |
| 6 | 78 | 78 | 83 | 100 | 85 |
| Avg. | 70 | 67 | 69 | 56 | 66 |

The detailed results for the different SRC and the average value for all uncoded and coded scenarios can be seen in Table 2 and in Fig. 2. The quantization step size of the individual percentage values is $(1/18) \approx 5.56$ because nine observers voted twice for each sequence. Overall, 70% of the observers preferred the 2D presentation over the 3D presentation. At the lowest quality, the rejection of 3D presentations becomes less pronounced, at the lowest bitrate

only 56% of the observers prefer 2D to 3D. A possible explanation for this effect is that the coding degradations mask the distortions by the depth induced blur.

The detailed analysis of the content shows that for one particular SRC, namely SRC5, most participants preferred the 3D presentation. The depth map of this sequence contains mostly values close to the central plane, thus the blurring effect is not obvious. At SRC2, the participants were undecided because this animation sequence contains a text moving out of the display. For about half of the observers, the strong blurriness due to the depth rendering of the text got compensated by the interest in the 3D sensation.

4.2. Experiment 2: Comparison between depth induced blur and resolution reduction

In the uncoded case, the 2D presentation was often preferred to the 3D presentation. The next step is to reduce the quality of the 2D presentation in order to achieve the same quality of experience as for the 3D presentation.

The texture of the 2D presentation is degraded by a resolution reduction process by introducing HRC-T1 that consists of a downsampling and upsampling process with a Lanczos-3 filter. As the downsampling factor increases, the 2D part of the screen exhibits more blur. An example for SRC1 can be seen in Fig. 3. Seven sequences were generated with 2D presentations using a downsampling factor from one to four in steps of 0.5 on the left side of the display and the undistorted 3D presentation on the right side of the display. The reversal of the sides led to another seven sequences.

Instead of using fourteen paired comparisons for each SRC, the observers used a different subjective test protocol, the “search for equal quality of experience”. The participant was asked to select the one out of the set of seven sequences which offered an identical quality of experience. Additional explanation was given that the selection should make it almost impossible to decide for a preferred sequence. The graphical user interface (GUI) that was used can be seen in Fig. 4.

It should be noted that the search for an identical sequence is limited by the available sequence set. The best quality for 2D presentation was the undistorted case with a downsampling factor of one, corresponding to the button “A” in Fig. 4. According to experiment 1, in about 30% the viewers still preferred the 3D presentation, thus no identical sequence could be selected. On the other hand, the worst 2D quality is downsampling with a factor of four, when the button “G” in Fig. 4 is selected. This sequence might still be preferred to the 3D presentation and again, no identical sequence could be selected. Thus, the GUI provided two more choices to indicate these two conditions.

The first advantage of using this “search for equal quality of experience” protocol is that the overall number of decisions in the subjective test is reduced from 84 to 12. The second advantage is that the full range of possibly identical sequences is available to the viewer at once so that he can carefully select the one that offers the same quality instead of having to remember his decisions. The drawback is that due to the fact that the available sequences only differ in texture quality, the observer may decide based on texture quality only instead of deciding based on his quality of experience.

The results are displayed in Table 3. In only one single vote of one observer for SRC4, the worst 2D quality was still preferred to the 3D presentation. In 23% the viewers preferred the 3D presentation even to the undistorted 2D presentation while in the paired comparison setup of experiment 1 about 30% of the observers preferred the 3D presentation. The difference is most prominent in SRC2 and SRC5



(a) Texture on the display



(b) Zoom on hand on left side of the display

(c) Zoom on hand on right side of the display



(d) Corresponding depth map

Fig. 3. Displayed content for a resolution reduction of 4 for SRC1

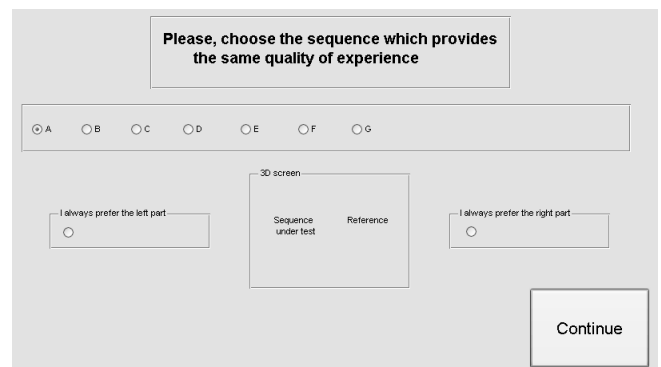


Fig. 4. Example of voting screen for question on identical quality

Table 3. Results for experiment 2a: Equivalence of downsampling 2D

| SRC | Factor | 3D preferred [%] | 2D always preferred [%] |
|------|--------|------------------|-------------------------|
| 1 | 2.5 | 6 | 0 |
| 2 | 2.3 | 33 | 0 |
| 3 | 2.1 | 11 | 0 |
| 4 | 2.2 | 11 | 6 |
| 5 | 1.7 | 61 | 0 |
| 6 | 2.7 | 17 | 0 |
| Avg. | 2.3 | 23.2 | 1 |

which were strongly preferred in 3D previously. This result may be an indication of the drawback mentioned above. One possible solution is to verify the results obtained from this subjective selection task on equal quality of experience with a paired comparison setup.

The average values in the second column of the table are calculated from the remaining results i.e., from those votes that indicated that a given sequence was equivalent in terms of quality of experience. The average downsampling factor is 2.3 with a minimum value of 1.7 for SRC5 and a maximum value of 2.7 for SRC6. On the considered display, this result would indicate that a 3D sequence with a texture and depth part of 960x540 pixels each is leading to the same average quality of experience as a 2D sequence of 417x243 pixels.

4.3. Experiment 3: Comparison between depth induced blur and coding artifacts

The experiment 3 is similar to experiment 2. Instead of resolution reduction, an H.264 codec was used to introduce degradations on the texture part. Again, several versions of 2D presentations were compared to one 3D presentation. Therefore, HRC-D1 was set to constant depth and HRC-D2 was not present.

The 3D presentation got distorted by a certain level of coding artifacts, thus HRC-T2 was set to H.264 coding at a specific QP_0 . The 2D presentation step HRC-T1 contained H.264 coding at several QPs that were generated relative to QP_0 , namely $QP_0+i \forall i \in \{-2, 0, \dots, 10\}$.

The subjective test protocol was exactly the same as in experiment 2, the subjects were asked to search for a sequence with equal quality of experience.

Table 4. Results for experiment 3: Equivalence of encoding at higher QP

| SRC | Change in QP | | | 3D preferred [%] | | |
|------|--------------|-----|-----|------------------|------|------|
| | $QP_0=29$ | 32 | 35 | 29 | 32 | 35 |
| 1 | 8.4 | 5.9 | 3.9 | 0 | 6 | 0 |
| 2 | 7.3 | 6.1 | 4.9 | 17 | 22 | 22 |
| 3 | 6.1 | 5.2 | 4.4 | 11 | 6 | 6 |
| 4 | 6.4 | 4.4 | 3.7 | 17 | 17 | 0 |
| 5 | 3.8 | 4 | 2.1 | 50 | 44 | 39 |
| 6 | 7 | 6.6 | 4.4 | 6 | 6 | 6 |
| Avg. | 6.5 | 5.4 | 3.9 | 16.8 | 16.8 | 12.2 |

The results are presented in detail in Table 4. In only about 17% the participants decided that the 3D presentation is always preferred. This value is lower than the 23% measured in experiment 2

and far lower than the 30% measured in experiment 1. Two reasons may be given. The first one is that all sequences were distorted, thus they may have preferred the 2D presentation in the distorted case. However, this contradicts to the results from experiment 1 when 3D was preferred to 2D more often at a higher QP. The second reason is that the subjects could vote for a better 2D texture quality than the 3D presentation had, namely QP_0-2 . The subjective test setup could be extended by more selections in the negative QP direction in order to reduce the usage of the preference button for 3D presentation further.

An equivalent quality of experience is reached on average if the QP for the 2D transmission is incremented by values in the range of 3.9 to 6.5. This would lead to a considerable reduction in the bitrate of the 2D sequence when compared to the 3D transmission. For SRC1, the change in QP is highest and the 3D version was preferred in only a single vote. For SRC5, the situation is reversed: Nearly 50% of the observers preferred 3D and the average QP change that was provided by the remaining observers is only about four.

The slight decrease from $i=6.5$ for $QP_0=29$ to $i=3.9$ for $QP_0=35$ may indicate that the difference between 2D presentations and 3D presentations gets smaller for larger distortions because the coding artifacts mask the depth rendering artifacts as was already mentioned in experiment 1. Another possible explanation is that the video quality degrades faster at higher QP values.

Based on the data gained from the last two experiments, the influence of the *degradation due to depth rendering* step in Fig. 1 on the quality of experience can be estimated. This is possible because an additional distortion step was added to the chain while the depth rendering was turned off. The depth rendering degradation can be related to a reduction in the resolution and to an offset in the QP of H.264 coding. However, the results strongly depend on the selection of the SRC sequences and there is a certain percentage of preference for the 3D presentation.

4.4. Experiment 4: Preference of limited depth sensation

In the previously described experiments, different versions of 2D images were compared to the reference 3D sequence. These experiments indicated that the results depend on the content and on the statistics of the depth map. In particular, for SRC5 many observers prefer the 3D version because the depth is centered around the central plane and the depth effect is moderate.

In order to test whether this positive effect could be applied to the other content as well, the depth maps of the 3D sequences were modified. In experiment 4, either no distortion was applied to the texture, thus HRC-T1 and HRC-T2 did not exist, or an equal H.264 coding distortion was used with a fixed QP of 32. One part of the display presented the 2D sequence, thus HRC-D1 was again set to constant depth. The other side of the display presented 3D with four different depth impressions, ranging from barely visible to very pronounced. In consequence, also the amount of depth induced blur changed from “noticeable but not annoying” to “very annoying”.

The four sequences were prepared by searching the minimum and maximum values of the complete video sequence. Then, the depth map values were linearly remapped to a new range which was centered at 128. The four sequences were generated with a deviation of ± 16 , ± 32 , ± 64 and ± 128 . Thus, the last sequence contained the maximum possible depth sensation ranging from 0 to 255. The logarithmic stepping was chosen based on the experience obtained from a pilot study.

A third subjective test protocol was used, the “search for the best quality of experience”. The observers were asked to choose the se-

quence which provided the best quality of experience. The 2D sequence on the other side of the screen was meant to provide a reference so that they could choose which 3D sequence they preferred most to the 2D version. However, in some cases, they still preferred the 2D presentation to all four 3D presentations. In the previous subjective test protocol, a concern was raised whether the subjects rated on the quality of experience scale because they select a matching sequence on a distortion scale. In this experiment, the selectable sequences differed in the amount of depth, thus it should be considered that the subjects may have voted on a mix of the “quality of experience” and the “added depth value” scale.

Table 5. Results for experiment 4: Choice of depth variation by observer

| SRC | Depth value range [128 ± x] | | 2D preferred [%] | |
|------|--------------------------------|------|---------------------|------|
| | uncoded | QP32 | uncoded | QP32 |
| 1 | 47 | 38.8 | 6 | 0 |
| 2 | 64 | 58.7 | 6 | 11 |
| 3 | 47.3 | 38.8 | 22 | 11 |
| 4 | 77.6 | 59.3 | 0 | 0 |
| 5 | 83.8 | 83.8 | 0 | 0 |
| 6 | 45.3 | 43.5 | 11 | 6 |
| Avg. | 60.8 | 53.8 | 7.5 | 4.7 |

The results are summarized in Table 5. In general, the observers preferred to have depth values which cover only half of the available 8 bit range, from 67 to 189.

For the natural content in SRC4 and SRC5, a far larger value was chosen, while for SRC6 the value was lower and a considerable number of people preferred the 2D presentation. For this sequence, the background got always blurred.

In SRC3, the observers also preferred the 2D sequence very often which was related to a status information overlay that is displayed at minimum depth, thus its text appeared to be blurred. Except for those two sequences, it seems that the experts found a depth setting where they preferred 3D over 2D. Thus, by using a limited depth the total number of 2D preferences was reduced from about 70% in Experiment 1 to about 7% for all sequences.

5. CONCLUSIONS

The first goal of this paper was to propose methodologies for subjective experiments that assess the quality of experience on autostereoscopic displays. The second goal was to identify and characterize the effect and impact of the depth rendering technology on autostereoscopic displays. Towards these goals experiments have been designed to compare the quality of experience of a 3D presentation to a 2D presentation.

Three different subjective test protocols were used. The first protocol was a “Pair Comparison” which can be expected to give the most precise results in terms of preference on the quality of experience scale. The second protocol can be termed “search for equal quality of experience”. The subjects are asked to select one sequence out of a set of sequences which were distorted along a certain axis of degradations. Two different axes were used, the reduction of the image resolution and the introduction of coding artifacts. Although the participants were asked to rate on the quality of experience scale, an indication was found that some participants may rate along the

provided axis of degradations. In this study, the degradations were only introduced in the texture part and the added value of depth may not have been taken fully into account. The third protocol uses the “search for best quality of experience”. The participants chose the best sequence out of a set of sequences with increasing 3D effect. The split-screen setup provided a 2D presentation as a reference so they could base their decision on the comparison of preference over 2D.

The results of the experiments led to several conclusions in line with our goal to characterize the effect of the depth rendering technology. The first experiment demonstrated that the rendering of the depth for a sequence introduced distortions in a way that about 70% observers preferred the 2D presentation without depth information on the same display. In the second experiment, the amount of this degradation was compared to a reduction of the resolution of the 2D presentation and it was shown that an equivalence is reached if the 2D sequence is downsampled by a factor of about 2.3 in both horizontal and vertical direction. In the third experiment, we evaluated the effect of the blur induced by the depth rendering on the perception of coding artifacts. The result is that a significantly higher bitrate may be necessary for the texture of 3D presentations in order to compensate for the depth rendering distortions.

These results differ between the various types of content included in the subjective test. Only one sequence out of the six sequences tested in experiment 1 provided a better quality when displayed in 3D. Starting from an analysis of this sequence it was concluded that a reduction of the depth map variation and moving the mean value of the depth map may lead to an improved 3D quality. In the fourth subjective experiment the viewers could choose the best 3D presentation from a set of four sequences. A setting for the depth range parameter was found for each sequence such that nearly all of the viewers preferred 3D to 2D presentation.

The next step would be to automatically adjust the depth range in order to maximize the quality of experience on the autostereoscopic display. Towards this goal, further investigations in the depth map rendering process and in the content dependency are necessary.

6. REFERENCES

- [1] Bruce E. Goldstein, *Sensation and Perception*, Wadsworth Publishing, 7 edition, 2006.
- [2] G. Leon, H. Kalva, and B. Furht, “3D Video Quality Evaluation with Depth Quality Variations,” *3DTV Conference: The True Vision – Capture, Transmission and Display of 3D Video*, 2008, pp. 301–304, 2008.
- [3] Question ITU-R 211/11, “ITU-R BT.500-10 Methodology for the subjective assessment of the quality of television pictures,” *ITU-R BT.500-10*, 1974.
- [4] ITU-T Study Group 12, “ITU-T P.910 Subjective video quality assessment methods for multimedia applications,” *ITU-T P.910*, 1997.
- [5] R. . G. Kaptein, A. Kuijsters, M. . T. . M. Lambooi, W. . A. IJsselstein, and I. Heynderickx, “Performance evaluation of 3D-TV systems,” in *Image Quality and System Performance V*, 2008, vol. 6808 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*.
- [6] Blender Foundation, “Big Buck Bunny Project Homepage,” <http://www.bigbuckbunny.org>.
- [7] ITU-T Study Group 16, “Advanced video coding for generic audiovisual services,” *ITU-T Recommendation H.264*, 2005.