

3D VIDEO QUALITY EVALUATION WITH DEPTH QUALITY VARIATIONS

Gustavo Leon, Hari Kalva, and Borko Furht

Dept. of Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL 33431
tavoadolfoleon@gmail.com, {hari, borko}@cse.fau.edu

ABSTRACT

In this paper we present the effect of depth quality on 3D video perception on autostereoscopic displays. This study was done using objective as well as subjective evaluation (using Phillips monitor). The goal of this work was to understand the impact of depth image quality and compression on the perceived 3D experience. Another objective is to evaluate the use of depth image quality as a measure of 3D quality. The experiments were conducted using a multi-view point Phillips 3D display that uses a single view and depth image. The displays are tested using subjective evaluation experiments. The results show that depth image can be compressed significantly, to 0.0007 bits per pixel, without affecting the 3D perception significantly. We found that there is a relation between the perception of depth, depth complexity and motion. Motion and complexity of the depth image have a strong influence on the acceptable depth quality in 3D videos. While depth quality gives useful information about 3D perception, additional work is necessary for developing objective metrics based on depth quality.

Index Terms— 3D video, depth, motion, quality evaluation

1. INTRODUCTION

Technologies for 3DTV are quickly becoming common. Recent standards activity in MPEG and the developments in the movie industry point to the increasing interest in 3D video technologies [1]. Movies use either polarized or active glasses in theaters for presenting 3D movies. Use of glasses is considered unnatural and cumbersome for home use and autostereoscopic displays are gaining attention. Developing metrics and evaluating the quality of 3D video is the subject of ongoing research but most of it was done without the use of autostereoscopic displays for their evaluation [2, 3]. We recently reported quality evaluation and asymmetric view coding on Sharp displays [4]. The new autostereoscopic displays from Phillips have nine view points and much easier to experience 3D.

The focus of our work has been on compression and quality evaluation for 3D video on autostereoscopic displays. The use of asymmetric view coding allows 3D

video services that use two views to reduce the total bandwidth required. Displays such as Phillips 3D display use a single view and a corresponding depth image and render the 3D views using the depth image based rendering. In this paper we report the impact of depth image quality on the 3D experience. One of the objectives of the work was to study the relationship between the depth quality and 3D experience. The depth quality was studied and compared with subjective evaluation of 3D experience.

The effect of depth quantization on quality was reported at the 3DTV Con in 2007 [5]. These experiments indicated that quantization up to 30 levels is sufficient. The impact of the content was not fully explored. One of the goals of this work is to use professionally prepared 3D videos with a range of complexity, motion, and depth to study the impact. The study was done using a 42 inch Phillips WOWvx autostereoscopic 3D display [6].

1.1 Video format in Phillips 3D displays

The input to the Phillips 3D Display is based on 2D plus depth (Z) format. Adjacent to the traditional 2D image a Z image is added, also called depth map. This is an image with the same size as the 2D image. Each pixel of the depth map has a corresponding pixel in the 2D image and indicates the distance of the corresponding 2D pixel to the observer.



Figure: 1 3D Image Format for Phillips monitor

The left half of the 3D video contains a 2D RGB picture, the right half contains a depth map in grey scale picture. The depth map values are interpreted as follows: 255 is close to camera / observer (= Max. value.), pixels with a depth value of 128 will be located on the screen and 0 represents far away, behind screen (= Min. value). The display resolution is 1920x1080 but the input video resolution is 1920x540; a 960x540 area on the left half has the 2D video and the right half has a 960x540 resolution depth image. A blank line inserted between two lines of the 1920x540 image to create

a 1920x1080 image for display. This 1920x1080 3D frame is applied to the DVI input of the Philips display. The display de-multiplexes the 2D and depth images and renders 9 different views.

The rest of the paper is organized as follows; Section 2 presents experimental setup and describes the content and tools used for evaluations. Results are presented in Section 4 and conclusions in Section 5.

2. EXPERIMENTAL SETUP

The quality experiments were conducted using a Philips 3D display and several 3D sequences with varying complexity, depth, and motion. The Philips 3D display is the 42 inch WOWvx display with 1920x1080 resolution.

2.2 Encoding depth map at different qualities

The effect of depth compression on video quality was tested by recompressing the depth image at various bitrates. The experiments were conducted using the 3D videos that were professionally created for the Philips monitor [7]. One shot from each of the six videos was selected for the experiments. Examples of shots used in the evaluation and a short description are shown in Figure 2 to 7.

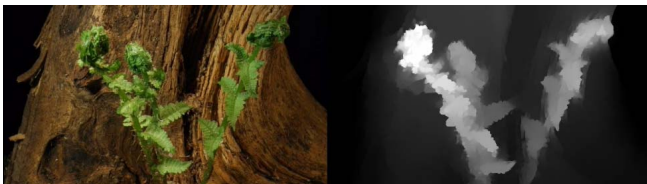


Figure 2. Shot 1: there is only one moving object but the leaves of the plant and the tree behind have multiple depths; complex depth image



Figure 3. Shot 2: there are two moving objects crossing each other and has many depth levels; complex depth image.



Figure 4. Shot 3: lot of movement in this video with multiple depth levels; complex depth image.

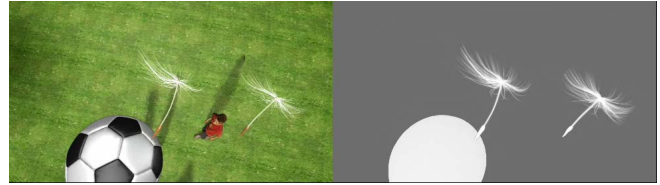


Figure 5. Shot 4: the background and foreground depth can be clearly seen with two values on the depth map; low complexity depth

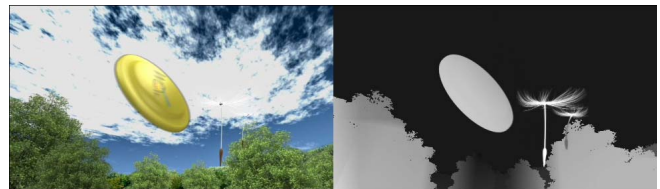


Figure 6. Shot 5: in this case the depth map has more than two values, but only has two moving objects, it is medium complexity

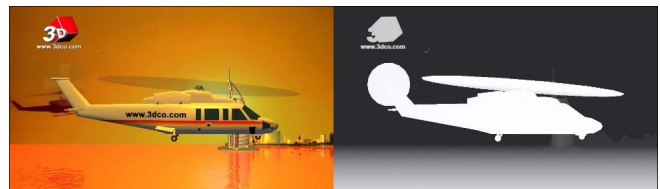


Figure 7. Shot 6: there is only one moving object and two depth levels; low complexity depth

The clips were selected to represent video with a range of complexity. These professionally created clips contain multiple objects, with varying depth, detail, and motion. Sequences include natural and synthetic videos. The complexity of the clips is determined by the number of objects, depth levels, and motion.

The process of recompressing depth is shown in Figure 8. A windows media decoder was used to decode the 3D sequence and the uncompressed video contains the 2D and depth images that stacked horizontally as shown in Figure 1. The depth information is separated and encoded using H.264 at various qualities. The compressed depth sequence is decoded and combined with 2D video create a composed 2D + depth uncompressed sequence with reduced quality of depth. The combined video is then encoded in WMV format using a Windows Media Encoder. To minimize the effects of WMV compression, the video is compressed at the best possible quality of 24 Mbps. The depth stream is encoded using H.264 at different bitrates; the bitrates range from 100kbps to 1000kbps in steps of 100kbps. For each shot, 10 different versions with varying quality are thus created. Subjective quality was evaluated with four depth qualities; depth encoded at 100, 300, 500, and 1000 Kbps.

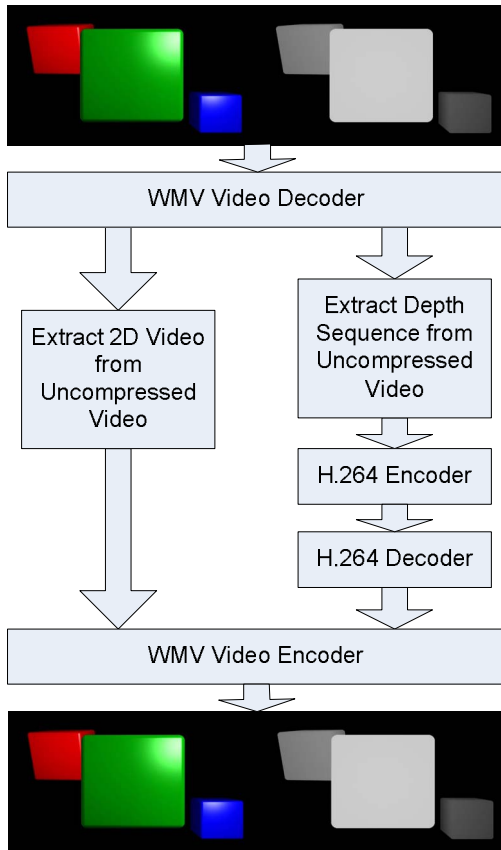


Figure 8: second experiment process

3. RESULTS AND DISCUSSION

Subjective evaluation was done using a set of 24 video clips. Each of the six shots discussed were encoded with four different depth qualities (100, 300, 500, and 1000 Kbps). These 24 video clips were shown to the users in a random order and the users were asked to rate the 3D experience from 1 to 5. Users were first presented with a set of original

quality 3D videos in order to provide them with the perfect 3D experience on the Philips displays. Users that could not experience 3D were not used in the evaluations. The results are based on 10 user evaluations. The users are graduate and undergraduate engineering students comprised of seven men and three women

3.2. Depth Image Quality

Depth image quality is expected to impact the quality of 3D views generated and hence affect 3D perception. PSNR is used as a measure of depth image quality. The quality of the depth images used in experiments is summarized in Table 1. Figure 9 shows the PSNR plots. As expected PSNR of the depth sequences drops with the bitrate. The complexity of the depth images is reflected in the RD curves. For the same bitrate, complex depth leads to lower PSNR. The helicopter sequence (shot 6) has the simplest depth and highest PSNR. Shots 1, 2, and 3 are complex and have the lower PSNR at a given bitrate. Shots 4 and 5 are in the medium range of complexity. Here complexity is defined as a subjective attribute that is a combination of motion, depth, and number of objects.

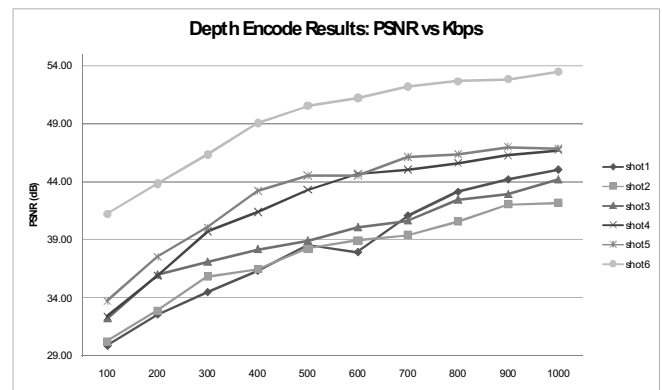


Figure 9: PSNR results from second experiment

Table 1: Depth quality

kbps	shot1		shot2		shot3		shot4		shot5		shot6	
	PSNR	BPP	PSNR	BPP	PSNR	BPP	PSNR	BPP	PSNR	BPP	PSNR	BPP
100	29.90	0.0007	30.25	0.0011	32.25	0.0008	32.43	0.0010	33.74	0.0010	41.23	0.0008
200	32.57	0.0013	32.88	0.0014	36.01	0.0014	35.92	0.0014	37.56	0.0014	43.84	0.0014
300	34.51	0.0020	35.85	0.0021	37.11	0.0021	39.74	0.0023	40.08	0.0020	46.32	0.0021
400	36.35	0.0027	36.49	0.0027	38.19	0.0028	41.41	0.0030	43.21	0.0027	49.03	0.0029
500	38.57	0.0035	38.27	0.0032	38.95	0.0034	43.30	0.0039	44.53	0.0034	50.54	0.0038
600	37.95	0.0042	38.95	0.0039	40.10	0.0042	44.69	0.0047	44.53	0.0043	51.21	0.0045
700	41.10	0.0050	39.43	0.0048	40.67	0.0050	45.05	0.0054	46.15	0.0051	52.16	0.0054
800	43.15	0.0058	40.60	0.0055	42.46	0.0059	45.63	0.0063	46.39	0.0059	52.65	0.0062
900	44.21	0.0066	42.04	0.0064	42.94	0.0066	46.28	0.0072	46.99	0.0067	52.83	0.0070
1000	45.05	0.0076	42.19	0.0073	44.20	0.0076	46.71	0.0080	46.88	0.0076	53.45	0.0079

For the subjective evaluation tests, having in mind that every evaluation can take a lot of time, we only use four bitrates, these are 100kbps, 300kbps, 500kbps and 1000kbps, the results are shown in Table 2 and Figure 10.

The mean opinion scores show that even with the depth sequence encoded at 100 Kbps, the perceived 3D quality is rated as very good for shots 4, 5, and 6 and rated as good for 1, 2, and 3. This can be attributed to the complexity of the video shots. When the 3D in a sequence is easily observable, the MOS is higher. For sequences with motion and varying depth, perceiving 3D is more difficult and depth image quality begins to affect the user experience quickly. Shot 2 has mixed MOS with 500 Kbps depth rated below 300 Kbps. This could be because of multiple objects and relatively high motion. Shot 3 is considered the most complex with people, cars, and buildings at various depths. As users' gaze moves among the objects in the scene, the depth perception suffers because of eye movement as users view point may change with the gaze on these types of displays. The complex scenes are thus more sensitive to depth quality compared to scenes with fewer objects of interest.

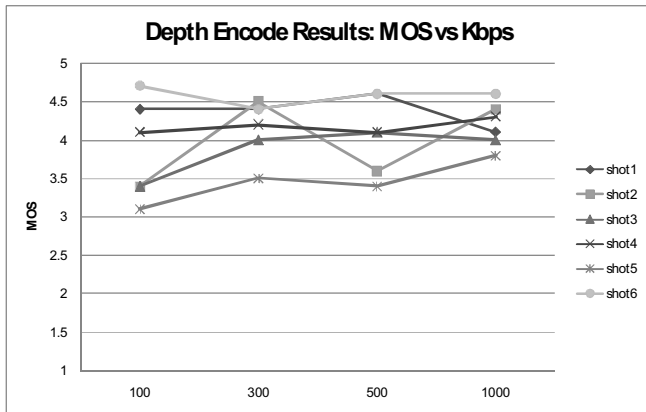


Figure 10: MOS results from second experiment

Table 1 summarizes the depth compression and quality. The results show that depth can be compressed by a large factor without affecting the quality significantly. The depth can be compressed by as much as 0.0007 bits per pixel without significantly affecting 3D perception. In general, higher PSNR, implying a higher depth quality, led to higher MOS. The same is not true at lower depth qualities as 3D experience in complex scenes tends to degrade quickly with PSNR. This is because of the motion characteristics and multiple depths of objects that are not reflected in PSNR. Motion in sequences has the strongest influence on 3D perception in autostereoscopic displays with multiple view points. Users' eye movements might lead to changes in view points and switching view points causes eye strain that probably results in a lower quality of user experience.

Table 2: MOS from subjective evaluation

kbps	MOS shot1	MOS shot2	MOS shot3	MOS shot4	MOS shot5	MOS shot6
100	4.4	3.4	3.4	4.1	3.1	4.7
300	4.4	4.5	4	4.2	3.5	4.4
500	4.6	3.6	4.1	4.1	3.4	4.6
1000	4.1	4.4	4	4.3	3.8	4.6

4. CONCLUSIONS

This paper presents the effects of depth image quality and compression on the 3D perception. The evaluation was performed on a 42 inch Philips autostereoscopic display with 9 view points. Evaluation was based on professionally created 3D video with varying depth complexity. The depth component of the 3D videos was extracted and compressed at various bitrates using H.264. 3D videos at various depth quality levels were subjectively evaluated. The videos used in the evaluation were selected to represent a range of complexity, motion, and depth. The results show that depth can be compressed significantly without affecting the quality of 3D experience. However, the complexity of a scene has a stronger influence at lower depth qualities. The nature of the multi-view point displays makes certain types of motion that could result in momentary switching of view point and this affects perceived quality. While the depth image can be compressed significantly, the complexity of the video and motion has to be taken into consideration in determining the acceptable levels of depth compression.

5. REFERENCES

- [1] H. Kalva, L. Christodoulou, L. Mayron, O. Marques, and B. Furht, "Challenges and opportunities in video coding for 3D TV," Special Session on "3-D TV: Primed for Success?," Proceedings of the IEEE International Conference on Multimedia & Expo (ICME) 2006, July, 2006, Toronto, Canada, pp. 1689-1692.
- [2] W. Ijsselsteijn and H. de Ridder and R. Hamberg and D. Bouwhuis and J. Freeman, "Perceived depth and the feeling of presence in 3DTV," Displays, 18 (1998), pp. 207-214.
- [3] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, and G.B. Akar, "Towards compound stereo-video quality metric: a specific encoder-based framework," IEEE Southwest Symposium on Image Analysis and Interpretation, 2006, vol., no., pp.218-222.
- [4] H. Kalva, L. Christodoulou, and B. Furht, "Evaluation of 3DTV Using Asymmetric View Coding Based on MPEG-2," Proceedings of the 3DTV Conference, Kos Island, Greece, 7-9 May 2007.
- [5] Ianir Ideses, Leonid Yaroslavsky, Itai Amit, Barak Fishbain, "Depth map quantization – how much is sufficient?," 3DTV-con, Tel-Aviv University, Israel, May 2007.
- [6] www.business-sites.philips.com/3dsolutions
- [7] <http://www.wowvx.com/video.html>