QUALITY ASSESSMENT OF 3D VIDEO IN RATE ALLOCATION EXPERIMENTS

A. Tikanmäki, A. Gotchev

Tampere University of Technology
Department of Signal Processing
Korkeakoulunkatu 10, 33720 Tampere, Finland

A. Smolic, K. Müller

Heinrich-Hertz-Institute (HHI)
Image Processing Department
Einsteinufer 37, 10587 Berlin, Germany

ABSTRACT

In this contribution, we address the problem of measuring and optimizing the visual quality of encoded 3D video. The 3D video is represented in the format of monoscopic color video augmented by per-pixel depth map and then encoded with H.264 encoder. To optimize the encoding performance we test different bit budgets for the color video and the depth and measure the quality by virtual view quality metrics. Small-scale subjective tests supplement the objective measurements. The obtained results show that, for similar overall quality numbers, observers favorably trade off lower depth quality for higher color quality and that depth distortions are perceived but considered less significant than the color distortions.

***Index Terms*—** 3DTV, H.264, video coding, video quality assessment, virtual view synthesis, video-plus-depth format

1. INTRODUCTION

In recent years, noticeable progress has been made in the development of autostereoscopic displays. This has created a widespread interest in designing and standardizing technologies needed for production, storage, delivery, and viewing of 3DTV content.

An efficient delivery format for 3DTV considers transmitting one view of color video augmented with a per-pixel depth map. The depth information can be used to render virtual views in which the objects of the monoscopic color video have been shifted to those positions where they would be seen from a virtual camera that is parallel to the real one.

Earlier studies [1] have shown that the depth map represented as a grayscale video can be compressed more efficiently than the corresponding color video, using as little as 10% to 20% of the color bit rate while retaining good quality. These results were obtained by subjective testing as there were no established objective video quality metrics (VQM) for 3D video.

In this paper, we address the problem of allocating bit budgets for color and depth information when both of them are compressed using H.264/AVC video codec. The rate-distortion characteristics of video plus depth content are estimated using a metric that is a straightforward extension of conventional monoscopic quality metrics to 3D. The performance of this metric, and the perception of quality in video plus depth content, is further analyzed by small-scale subjective testing.

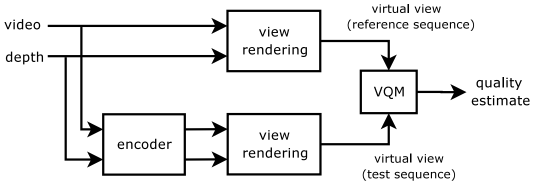


Fig. 1. Virtual view video quality metric

2. VIRTUAL VIEW VIDEO QUALITY METRIC

A conventional monoscopic VQM can be used as a quality measure for video plus depth content by measuring the quality of the virtual views that are rendered from the distorted color and depth sequences. The undistorted reference sequence that is needed for the full reference VQM is obtained by rendering virtual views from the original color and depth maps. This process is illustrated in Figure 1.

The virtual views are rendered by shifting the objects in the color image to those positions where they would be seen at when looking from a virtual camera that is parallel to the real one. Because the new positions are calculated from the depth map, the virtual view synthesis fuses together the information from both the depth and the color sequences, and hence we can assume that the quality of the virtual view somehow represents the quality of both component sequences.

Practically any VQM could be used for comparing the virtual views. In this study we use VSSIM, which is a version of Structural SIMilarity metric (SSIM) adapted for video [2]. The performance of VSSIM is compared to that of PSNR.

SSIM measures the local structural similarity of two images by eliminating the effect of contrast and luminance differences. VSSIM, the video version of the metric, analyzes only a fixed number of blocks randomly sampled from each video frame in order to reduce the computational requirements. It also uses the amount of motion in the scene as weights when the local SSIM index values are pooled to form the overall quality score.

3. VIDEO PLUS DEPTH RATE ALLOCATION

The virtual view VQM introduced in Section 2 was used to estimate how the overall quality of video plus depth content depends on the bit rates used for encoding the depth map and color video.

Four different source sequences were used as test material. Two of them, “Interview” and “Orbi” (720x576, 25fps), were from the ATTEST project, and the two others, “Ballet” and “Breakdancers” (1024x768, 15fps), were single views chosen from the multi-view datasets produced by Microsoft Research. The depth maps of the ATTEST sequences had lower contrast but more detail than the Microsoft sequences.

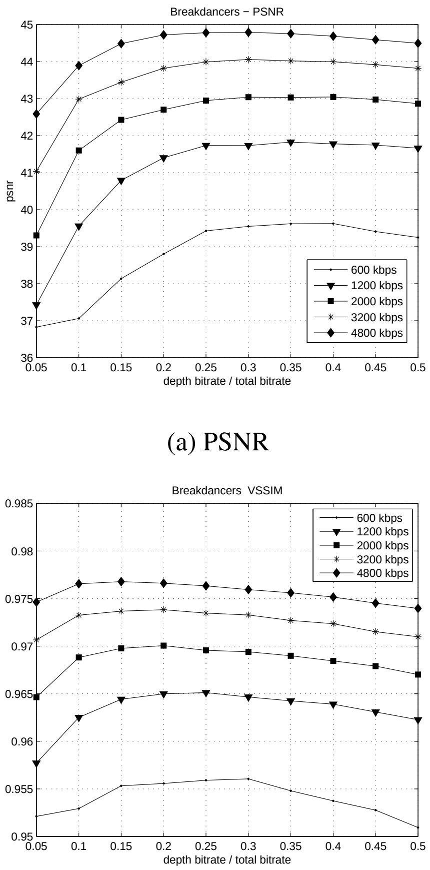
The test sequences were encoded with different total bit rates ranging from 600 kbps to 4800 kbps. Within each total bit rate, we encoded the depth sequence with bit rates ranging from 5% to 50% of the total bit rate. The rest of the total bit rate was used for encoding the corresponding color sequence. Both color and depth were encoded to Baseline H.264 using the same encoder settings. The virtual views were rendered using the approach described in [3].

Figure 2 shows the quality of the test sequence “Breakdancers” measured with virtual view PSNR and VSSIM. These graphs have one curve for each total bit rate, and the x-axis values from left to right correspond to depth bit rates ranging from 5% to 50% of the total bit rate.

From the curves it can be seen that the best quality in terms of PSNR is achieved by spending 30–40% of the total bit rate for depth coding. With VSSIM the quality peak is at lower depth bit rates, around 15–30%. The bit budget division yielding best quality tends to require proportionally lower depth bit rate when the total bit rate increases because the quality of the depth sequence saturates faster than the color quality.

The two metrics have characteristically different rate-distortion curves. With PSNR, the quality drops quite rapidly when the proportion of bit rate used for depth map decreases below 25% but the PSNR differences between the 25% and 50% points are practically negligible. The VSSIM curves, on the other hand, show a clear drop only below 5% depth bit rates at the low end, and a steady degradation of quality when moving from the optimal bit rate division towards the 50% point.

The rate-distortion curves of the other test sequences are quite similar as the ones presented here. The sequences are somewhat less demanding from the depth coding point of view and hence the VSSIM quality peak is achieved at slightly lower depth bit rates (10–20% of total bit rate). The PSNR curves are similar as well, showing a rather flat quality peak around the 30% point. The “Ballet” test sequence was something of an exception as its PSNR peaked at the 50% point.



(a) PSNR

(a) VSSIM

Fig. 2. Virtual view quality of “Breakdancers”

The virtual view PSNR appears to be rather unstable measure when the depth maps are distorted. The left and right virtual view PSNRs of the same sequence could differ by up to 2dB while the VSSIM of different views stays virtually the same. PSNR is also very sensitive to distortions in the depth map as there are, generally speaking, no significant PSNR differences between using 25% and 50% of the data for depth coding. VSSIM, on the other hand, shows clear improvement of quality when color bit rate increases at the expense of depth bit rate.

4. SUBJECTIVE TESTS

A small-scale subjective evaluation session was organized in order to establish an understanding of the relation between objective quality estimates and actual human perception. The test was conducted by displaying test sequence pairs to test participants and asking them to choose the better quality sequence of each pair. No quantitative quality grades were col-

	Sequence A	bit rate	Votes A	Unsure	PSNR A	VSSIM A
	Sequence B	color–depth	Votes B		PSNR B	VSSIM B
1	**A**: Breakdancers	2660–140	8	2	40.27	0.970
	B: Breakdancers	440–360	0		40.44	0.958
2	**A**: Breakdancers	960–240	1	2	41.24	0.965
	B: Breakdancers	1900–100	7		38.84	0.965
3	**A**: Breakdancers	1900–800	7	2	43.90	0.972
	B: Breakdancers	1900–100	1		38.84	0.965
4	**A**: Breakdancers	440–360	0	1	40.44	0.958
	B: Breakdancers	640–160	9		39.90	0.961
5	**A**: Ballet	250–150	3	2	42.05	0.973
	B: Ballet	1330–70	5		42.35	0.982
6	**A**: Ballet	300–300	4	0	42.94	0.975
	B: Ballet	550–30	6		39.10	0.980
7	**A**: Ballet	1330–70	3	4	42.35	0.982
	B: Ballet	550–450	3		45.06	0.982
8	**A**: Ballet	550–30	2	3	39.10	0.979
	B: Ballet	550–450	5		45.06	0.982
9	**A**: Interview	330–330	0	1	38.50	0.953
	B: Interview	630–30	9		38.17	0.964
10	**A**: Interview	500–530	2	2	40.61	0.967
	B: Interview	600–30	6		40.18	0.968
11	**A**: Interview	900–30	6	1	38.65	0.970
	B: Interview	500–530	3		40.37	0.967

Table 1. Votes of the subjective evaluation

lected. The quality information obtained in the test was then compared to the objective quality values given by the virtual view metrics.

Altogether 10 people participated in the evaluation. Most of the participants were young students of information technology but they had no previous experience of autostereoscopic 3DTV. Their opinions on 3D video quality may differ from those of people accustomed to this technology. The test sequences were made of three source videos: "Breakdancers", "Ballet", and "Interview". They had varying amounts of distortions caused by H.264 compression in the color and depth components. The test consisted of altogether 11 test sequence pairs. The test sequences were displayed using Heinrich-Hertz Institut's *Free2C*, a lenticular autostereoscopic display equipped with automatic eye-tracking that was also used in [1].

Table 1 summarizes the data collected in the evaluation sessions together with the objective quality estimates obtained using the virtual view metric. It can be seen that there is no clear viewer preference for either of the sequences in test pairs 5–7. Apparently "Ballet" sequence is not well suited for observing quality differences and therefore the following analysis of results will mainly concentrate on "Breakdancers" and "Interview".

When the viewer preferences are compared with the objective quality estimates, we notice that when the quality values are similar, the viewers always prefer the sequence with higher color quality and lower depth quality.

The test pairs 1 and 9 have test sequences with almost the same PSNR, while the VSSIM scores prefer the sequences with higher color bit rates and hence higher color quality. The test pairs 2, 4, 10, and 11 have sequences with almost the same VSSIM qualities. For these test pairs, PSNR gives at least slight preference for the sequences with higher depth quality. In both of these groups, the majority of test participants considered the sequences having higher color quality, and thus lower depth quality, to be better.

Test pairs 3 and 8 were included to test the participants' depth perception. The sequences of these pairs had the same color sequence while the depth sequences had a considerable quality difference. The majority of the viewers preferred the sequence having higher depth quality, which implies that they were able to notice the depth distortions but considered them to be less significant than the color distortions.

5. CONCLUSIONS

This paper studied how the quality of 3D videos stored as a monoscopic color video augmented by a per-pixel depth map is affected by the bit budgets used for coding the color and depth data. Both the color and the depth were compressed using Baseline H.264. The video quality was estimated using PSNR and VSSIM of the virtual views that were synthesized from the compressed color and depth sequences. These objective quality measurements were further assessed in a small subjective quality test.

The rate division-distortion curves of VSSIM showed that highest quality can be achieved when 15–20% of the total bit budget is used for coding the depth map. PSNR favored higher depth quality than VSSIM. For most test sequences, it showed that quality remains almost the same at depth bit rates between 25–50% of the total, while below 25% the quality drops steeply. Of these two metrics, VSSIM was closer to the subjective evaluation results.

The subjective evaluation suggested that human viewers consider depth distortions to be less important than color distortions in situations where the objective quality metrics give equal grades. This shows that the quality estimation method is not perfect but better metrics are needed. Although VSSIM of the virtual view is not perfect quality estimation method, it gives a conservative upper bound for the bit rate needed for coding the depth map. Larger subjective tests with more varied test material are needed to establish better understanding of 3D video quality and to develop more accurate methods for measuring the quality.

6. ACKNOWLEDGMENTS

This work is supported by EC within FP7 under Grant 216503 with the acronym MOBILE3DTV and by the Academy of Finland, project No. 213462 (Finnish Centre of Excellence program (2006 - 2011).

7. REFERENCES

[1] C. Fehn, K. Hopf, and B. Quante, “Key technologies for an advanced 3D TV system,” *Proceedings of SPIE*, vol. 5599, pp. 66, 2004.

[2] Z. Wang, L. Lu, and A.C. Bovik, “Video quality assessment based on structural distortion measurement,” *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121–132, 2004.

[3] C. Fehn, “Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3 D-TV,” *Proceedings of SPIE*, vol. 5291, pp. 93–104, 2004.