OBJECTIVE QUALITY MEASURES COMPARISON OF IMPAIRED 3D VIDEO SEQUENCES FROM THE UC3D DATABASE

[1]Emil Dumic, [1]Sonja Grgic, [2,3]Luis Alberto da Silva Cruz, [2,4]Pedro Assuncao

[1] Department of Wireless Communications,
Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia
[2] Instituto de Telecomunicações,
[3] Dep. of Electrical and Computer Engineering, University of Coimbra, Portugal
[4] Instituto Politecnico de Leiria - ESTG, Leiria, Portugal

ABSTRACT

In this study we analyze the correlation between different image and video quality measures and corresponding subjective scores computed on 20 3D texture+depth video sequences from the University of Coimbra 3D (UC3D) video database. Quality metrics were tested on (degraded) depth information and synthesized views. The results show that in the case of measures based on depth information, the best correlation is achieved by SSIM. For the case of measures computed on the synthesized view, the highest correlation is obtained when using PSNR. Overall results show that quality measures computed on the synthesized view correlate better with subjective (DMOS) scores than measures based on the depth information.

Index Terms — Objective quality measures, subjective scores, correlation, 3D video sequences

1. INTRODUCTION

Quality of media content can be evaluated using different types of measures. Generally, quality can be evaluated using objective and subjective scores. Subjective scores can be obtained using visual experiments conducted under controlled conditions, in which human observers grade the quality of the multimedia contents [1]. Such experiments give the most precise scores (MOS, Mean Opinion Score), but are usually time consuming and expensive. Because of that, objective measures have been developed to mathematically estimate human opinion about visual quality. Every objective quality measure has as its aim approximating the human perception of quality as closely as possible, meaning that good correlation with subjective measures (Mean Opinion Score, MOS) is sought [2].

Objective quality measures for image and video can be generally divided into three categories according to the reference information they use, as follows:

- full-reference (FR) quality measures;
- reduced-reference (RR) quality measures;
- no-reference (NR) quality measures.

This paper presents an empirical study on the performance of existing FR and RR 2D image and video quality measures applied to 3D video sequences represented in texture+depth format which have been subjected to a specific degradation, namely depth map corruption due to the packet losses during transmission.

The paper is organized as follows: section 2 describes related work, section 3 introduces the University of Coimbra 3D (UC3D) video dataset, section 4 describes some of the existing FR and RR 2D image and video quality measures and their application to the UC3D dataset, section 5 presents experimental results and section 6 draws the conclusions.

2. RELATED WORK

3D video has recently become a format of foreseeable interest for visual information diffusion. For this reason the problem of evaluating the end-user 3D video perceived quality started attracting the attention of video coding and transmission experts in academia and industry.

A few approaches to 3D video quality evaluation have their roots in similar methods developed for 2D media. Some of those quality evaluation procedures and algorithms are surveyed in [3] that present a classification and comparison of current 2D objective image and video quality assessment algorithms, evaluated on different datasets of subjectively graded 2D video contents.

In [4] Joveluro et al. proposed the use of full-reference 2D video Perceptual Quality Metric for 3D video quality evaluation. This metric is based on quantifying the brightness and contrast distortions of the luminance component of synthesized views. The distortions magnitudes are estimated by variances of the brightness and contrast weighted by the mean of each pixel block. A dataset of subjective evaluations was assembled based on texture+depth video encoded with scalability using Joint Scalable Video Model (JSVM) at different quality factors. The quality of synthesized views was evaluated using different 2D image and video quality metrics.

In [5] Hewage et al. describe a reduced-reference quality measure based on feature extraction from structural degradation of depth maps in 3D texture+depth video sent over video transmission network subject to different packet loss rates. Depth maps of tested sequences were encoded using the H.264/AVC video coding standard (JM reference software Version 16.0) with different quality factors. Performance of the method shows accurate results compared to its counterpart Full-Reference (FR) quality metric.

In [6] Bosc et al. investigates the reliability of objective quality metrics commonly used for the quality assessment of 2D visual media, in the context of 3D texture+depth video. Two approaches for new full-reference 3D video quality measure were proposed: the first one is based on the analysis of the shifts of the contours of the synthesized view; the second one is based on the computation of a mean SSIM score of the disoccluded areas.

978-1-4799-4758-4/14/$31.00 ©2014 IEEE

3. UC3D DATABASE

The UC3D database consists of four different original 3D video sequences in texture+depth format with 10 seconds of duration. For each original sequence, five degraded sequences were produced with different frame burst lengths and different number of affected frames. In the set of four sequences, three of them have 300 frames (30 fps) and the other has 250 frames (25 fps). The degradations introduced to generate the impaired videos affect only the depth maps and simulate losses of packets transporting encoded depth data [7]. For each impaired depth map sequence a (new, degraded) synthesized view was rendered using the original texture information (and degraded depth map). To allow quality comparisons leading to the grading of the effect of depth information losses, for each of the four source sequences a non-degraded view was synthesized, based on the original texture and original depth information.

Details about each of the tested 3D video sequences are presented in Table 1.

Table 1 - Original 3D Sequences

Sequence	Resolution	Frame Rate	Frames
Balloons	1024x768	30fps	300
Kendo	1024x768	30fps	300
Champagne	1280x960	30fps	300
poznancarpark	1920x1088	25fps	250

Subjective evaluations were made using DSCQS (Double Stimulus Continuous Quality Scale), which conforms with ITU-R BT.500-11 [8]. The test media consisting of original and degraded video sequences were presented to 35 observers (aged 21 to 47 years), whose visual and depth acuity was verified using the "Fly" test. The 3D visual stimuli were displayed on a 20 inch autostereoscopic 3D Philips WOWvx (9-view) display. Monitor settings used were the factory defaults. To obtain a larger result set in a shorter time and to prevent eye fatigue due to 3D content viewing, each reference-distorted pair was presented twice (in random order). The viewers knew beforehand that in all the pairs the first video to appear was the reference. In this way, 40 scores per observer were recorded. For each observer a weighted mean of the 2 grades pertaining to the same reference-distorted pair was calculated yielding a (differential) quality score [1]. At the end, 20 difference mean opinion scores (DMOS) were calculated.

For each affected video sequence, a number of key statistical parameters describing the extent and distribution of the degradations were computed, such as the average number of frames affected by the depth losses (affected frame rate - AFR) and the length and frequency of bursts of affected frames.

Figure 1 shows the dependency between AFR and DMOS scores. It can be seen that AFR is already well correlated with DMOS scores (Spearman's correlation is 0.9023), which probably increases correlation with all objective measures.

Figure 2 shows histograms of the length of bursts of affected frames for the best and worst case of balloons video sequence is shown in Figure 2. This type of information (burst length and respective frequencies) could be potentially used in the future work to support the proposal of a no-reference objective quality measure.

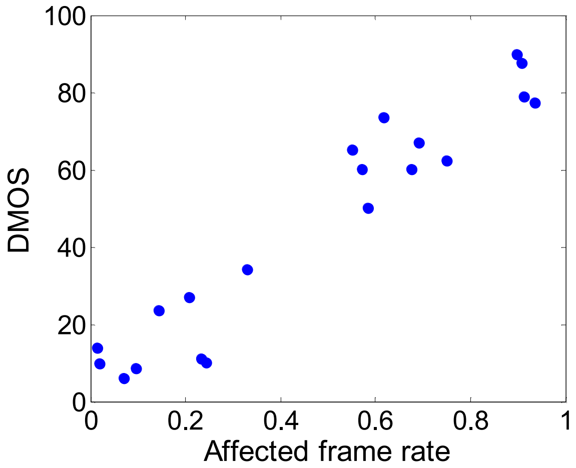
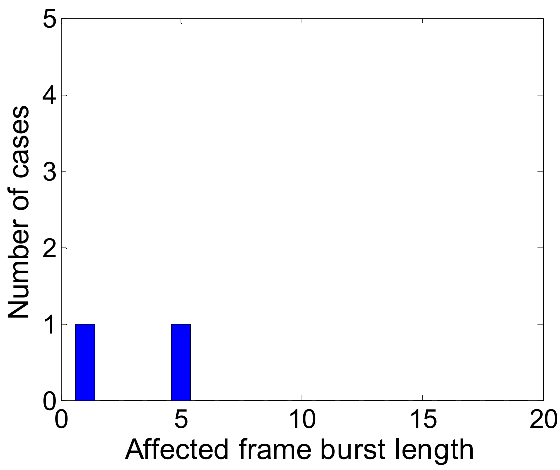
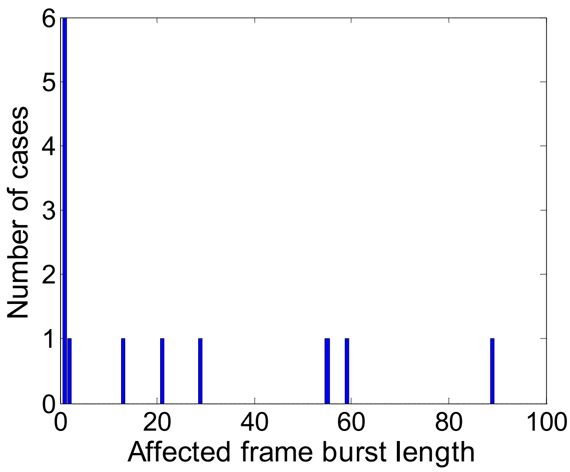


Figure 1. Affected frame rate versus DMOS



a)



b)

Figure 2. Histogram of affected frame length versus number of cases, UC3D database, video 1 (balloons): a) best case (AFR= 2%), b) worst case (AFR= 91.33%)

4. EVALUATED QUALITY MEASURES

A total of 4 original and 20 degraded 3D video sequences were evaluated through the application of objective (i.e. computable) quality measures, using either the depth information or the synthesized view information.

Several full reference image and reduced reference 2D video measures were computed and the scores obtained were compared with the DMOS (Difference Mean Opinion Score) values of the previously described 3D database.

The measures used were:

- MS-SSIM (Multiscale Structural Similarity Index) [9],
- PSNR (Peak Signal to Noise Ratio),
- SSIM (Structural Similarity Index) [10],
- VIF (Visual Information Fidelity) [11],
- VSNR (Visual Signal-to-Noise Ratio) [12],
- IW-SSIM (Information Weighted SSIM) [13],
- IW-PSNR (Information Weighted PSNR) [13],
- IQM2 (Image Quality Measure 2) [14] and
- RVQM (Reduced Video Quality Measure) [15].

RVQM is a 2D video quality measure, while the others are image quality measures.

IQM2 was calculated using SPWT (Steerable Pyramid Wavelet Transform) with 2 orientations as recommended [14]. On each scale and orientation of SPWT transform a modified SSIM (only structure and contrast component) was applied and final measure was calculated by multiplying the modified SSIM grades over all subbands.

RVQM [15] was calculated using 272 (width) x 272 (height) x 32 (frame) pixels, step size 16 pixels, 1st order and 3rd component of the Riesz transform. On each scale modified SSIM (only structure and contrast component) in 3 dimensions was applied and final measure was calculated by multiplying modified SSIM grades over all subbands.

For each sequence a quality score was computed for the corresponding depth map sequence and a second score was computed for the respective synthesized view sequence. In the case of image-based quality measures, the video-level quality scores were calculated as the average of the individual frame-level quality grades of all frames in the sequence. The original resolution of the test sequences (given in Table 1) were provided to each tested metric.

It should be noted that, because of the different post-processing techniques that WOWVx display uses, to render its own views, it is possible that tested metrics did not include all effects on final video quality. In this experiment, the monitor used all of its default values.

5. RESULTS

The quality estimates produced by the previously described image and video quality measures have been compared to the database subjective scores using Spearman's rank order correlation. This correlation measure assesses how well an arbitrary monotonic function can describe the relationship between two variables, without making any assumptions about the frequency distribution of the variables. Spearman's correlation coefficient is calculated like Pearson's correlation but over ranked variables. Overall results are presented in

Table 2 and Figure 3 from which it can be concluded that among all the tested measures, the best results were obtained using SSIM of the depth maps and PSNR of the synthesized views. RVQM measure gives second best results in both cases.

For the measures applied to the depth information, SSIM, PSNR and RVQM have Spearman's correlation higher than 0.9, while in the case of synthesized view-based quality scores, all measures beside VSNR have correlation higher than 0.9. Only VSNR showed lower correlation in both depth and synthesized measure comparisons. This is probably due to the VSNR calculation in logarithmic scale (dB) and, in some cases, the occurrence of many undistorted frames per video sequence, in which cases VSNR gives high but finite score. Comparison using depth view was made because it can point out future directions towards new 3D video quality measures. For example, reduced reference measure using only finite number of parameters from depth view or no reference measure with parameters extracted from degraded bitstream could correlate well with HVS [7].

When comparing results from depth and synthesized views, nearly all measures have better correlation in synthesized view (only SSIM and VSNR have lower correlation in synthesized view). This is probably because synthesized view carries more visual information than depth, so it is better correlated with HVS.

As described earlier, AFR is well correlated with DMOS scores, which could have impact on the final correlation between tested objective measures and DMOS (by increasing it). In the future, the use of information about this type of degradation (depth map degradation due to the packet losses) to estimate the synthesized view quality could be more thoroughly evaluated using more test video sequences with equal AFR but with different affected frame burst length.

Table 2 - Spearman's correlation between different metrics and DMOS, UC3D database

	MS-SSIM	PSNR	SSIM	VIF	VSNR	IW-SSIM	IW-PSNR	IQM2	RVQM
Spearman's corr depth	0.8827	0.9293	**0.9474**	0.8842	0.6722	0.8902	0.8992	0.8947	0.9293
Spearman's corr synth	0.9368	**0.9639**	0.9338	0.9278	0.4571	0.9564	0.9564	0.9383	0.9579

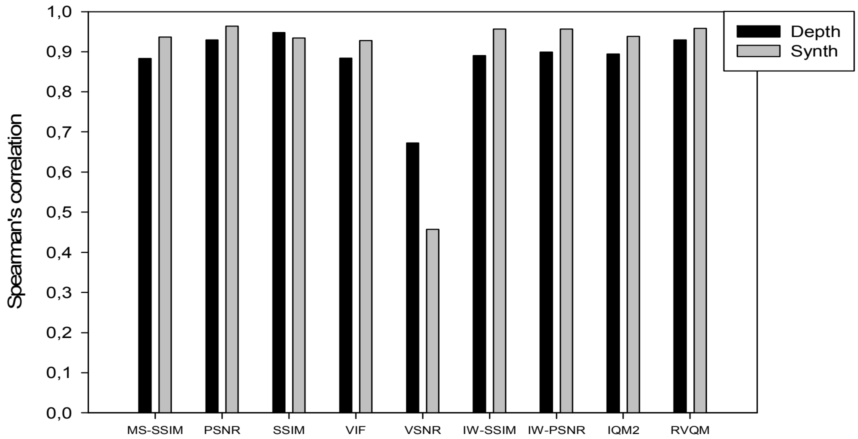


Figure 3. Spearman's correlation between different quality metrics and DMOS, UC3D database

6. CONCLUSIONS

In this paper we tested existing 2D objective image and video quality measures on a new 3D video (texture+depth) subjective quality database, the University of Coimbra 3D (UC3D) database. From the overall correlation results presented and discussed in the previous section, we conclude readily that from among all the measures tested, the best results were obtained using SSIM measure for depth map comparison and PSNR for synthesized view comparison. RVQM gave second best results in both cases. Also, in nearly all tested measures better results were obtained using synthesized view than using depth information, probably because the former carries visual information that is better correlated with the information extracted and used by the quality perception mechanisms embedded in the human visual system.

Future research could include development of RR and NR objective measure based on AFR and some other properties of tested 3D video sequences, like spatial and temporal information or some other features. AFR, as well as affected frame burst length, could potentially be used as a good starting point, because it showed good correlation properties with DMOS, Figure 1.

Development of new 3D database using different degradations seem promising, both induced by compression and processing operations as well as transmission impairment (like packet losses and streaming freezes).

7. ACKNOWLEDGMENT

The work described in this paper was conducted under the research project EU COST Action IC1105 (3D-ConTourNet).

8. REFERENCES

[1] ITU-R BT.500-11 "Methodology for the subjective assessment of the quality of television pictures", International Telecommunication Union/ITU Radiocommunication Sector, 2002

[2] Video Quality Experts Group, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Multimedia Quality", http://www.vqeg.org/, 2008

[3] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison", IEEE Transactions on Broadcasting, vol. 57, no. 2, pp. 165-182, Feb. 2011

[4] P. Joveluro, H. Malekmohamadi, W. A. C. Fernando, and A. M. Kondoz, "Perceptual Video Quality Metric for 3D video quality assessment", 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2010, pp. 1–4, 2010

[5] Hewage, C.T.E.R., Martini, M.G., " Quality evaluation for real-time 3D video services", IEEE International Conference on Multimedia and Expo (ICME), pp. 1-5, 2011

[6] E. Bosc, R. Pepion, Patrick Le Callet, M. Pressigout, L. Morin, " Towards a New Quality Metric for 3-D Synthesized View Assessment ", IEEE Journal of Selected Topics in Signal Processing, Vol 5 , No. 7, pp. 1332-1343, 2011

[7] R. S. Soares, "Quality Evaluation of 3D Video Subject to Transmission Channel Impairments", master thesis, 2013

[8] Recommendation ITU-R BT.500-11: "Methodology for the subjective assessment of the quality of television pictures", 2002

[9] Z. Wang, E.P. Simoncelli, A.C. Bovik, "Multiscale structural similarity for image quality assessment", 37th Proc. IEEE Asilomar Conf. on Signals, Systems and Computers, Vol. 2, pp. 1398-1402., 2003

[10] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Trans. on Image Proc., Vol. 13, No. 4, pp. 600-612., 2004

[11] H.R. Sheikh and A.C. Bovik, "Image information and visual quality", IEEE Trans. Image Processing, Vol. 15, No. 2, pp. 430-444., 2006

[12] D.M. Chandler and S.S. Hemami, "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images", IEEE Transactions on Image Processing, Vol. 16, No. 9, pp. 2284-2298., 2007

[13] Z. Wang, Q. Li, "Information Content Weighting for Perceptual Image Quality Assessment", IEEE Trans. on Image Processing, Vol. 20, No. 5, pp. 1185-1198., 2011

[14] E. Dumic, S. Grgic and M. Grgic, "IQM2 - New image quality measure based on steerable pyramid wavelet transform and structural similarity index", Signal Image and Video Processing, accepted for publication

[15] E. Dumic, S. Grgic, "Reduced Video Quality Measure Based on 3D Steerable Wavelet Transform and Modified Structural Similarity Index", Proc. of the 55th International Symposium ELMAR-2013, pp. 65-69, 2013