OBJECTIVE BITSTREAM QUALITY METRIC FOR 3D-HD VIDEO

Dawid Juszka, Lucjan Janowski, Zdzisław Papir

AGH University of Science and Technology
Department of Telecommunications
al. A. Mickiewicza 30
30-059 Kraków, Poland

ABSTRACT

Now growing interest in 3D video services inevitably brings a greater need to measure their quality. From a financial point of view alone, it is crucial for providers to offer users the best experience possible. Such measurements are necessary in order to optimize the performance of telecommunication networks or to plan new investment. Unfortunately, the nature of this kind of service makes it almost impossible to use existing and verified 2D quality metrics. Therefore, suitable 3D quality metrics are urgently needed. In this paper, we propose one such solution - an objective, bitstream quality metric for stereoscopic high definition video affected by compression and packet loss in a network.

Index Terms— objective metric, bitstream metric, 3D, HD, QoE, Quality of Experience

1. INTRODUCTION

In order to generate income and remain competitive, telecoms operators need to calibrate their networks using a human-centric approach. A satisfying quality of experience (QoE) is the *sine qua non* for customers who use services offered by providers. For a number of years, scientists from around the world have been conducting research into measuring this experience for multimedia services. Frequently, the results of such research have been quality metrics - functions providing a MOS (Mean Opinion Score) prediction depending on the chosen parameters of the content delivery chain. These functions are built on statistical analysis of subjective evaluation results in relation to objective parameters. As such, they support a human-centric paradigm and help operators to meet customers' needs. They are easily employed to predict the level of end-user satisfaction with the offered video quality. Using this knowledge, proper network parameters can be readjusted for better quality performance, which in turn helps to guarantee commercial success in a competitive market.

The most novel multimedia services currently emerging are those connected with 3D video distribution. This kind of multimedia enables users to perceive an effect of depth in a way which is impossible to achieve by methods deployed in 2D technology - such as occlusion, shadowing or geometrical perspective. Potential applications for three dimensional projections include education, entertainment, surgery, video conferencing and many more.

It seems likely that take-up for 3D content will increase considerably in the near future. One basis for such conclusion is the Cisco Visual Networking Index: Forecast and Methodology [4]. This foresees that between 2011 and 2016, the compound annual growth rate (CAGR) for traffic generated by 3D Video-on-Demand (VoD) will be equal to 109%. This means that Global IP Traffic for this service will grow from 2 petabytes per month in 2011 to 82 petabytes per month in 2016. A corresponding conclusion can be derived from the "3D display technology and market forecast report" delivered by NPD DisplaySearch [3]. According to this source, the total stereoscopic 3D display market will grow from 0.7 million units and $902 million in revenues in 2008 to 196 million units and $22 billion in revenues in 2018, a compound annual growth rate (CAGR) of 38% for revenues and 75% for units. This data justifies the interest of operators in metrics of QoE for 3D video.

This paper presents a bitstream metric which enables 3D quality measurement in real time without full decoding. It takes into account the influence of compression and packet loss on the quality of stereoscopic FullHD video presented using an active shutter display. To build this metric, results and conclusions from a subjective QoE experiment were used.

The reminder of this paper is as follows: in Section 2, research related to the state of the art is provided. In Section 3, the main elements and features of the experiment are described, including the origin and modifications of the test material. A more detailed presentation of the metric, its components, and the method of deriving QoE from a MOS scale, are described in Section 4. Conclusions and future work are presented in Section 5.

This work has been supported by Deans Grant (agreement number 15.11.230.092) - Faculty of Computer Science, Electronics and Telecommunications, AGH University of Science and Technology.

2. RELATED WORK

Many scientific teams find research on Quality of Experience for 3D video inspiring and challenging. This has led to proposals for some objective metrics, each one different in terms of design approach and usage scenario. The usefulness of 2D image quality metrics for 3D images has been evaluated in several studies [12, 5, 8, 9]. Campisi et al. compared several 2D metrics for stereoscopic images and concluded that, while their performance is state-of-the-art with regard to 2D images, they are not satisfactory for 3D imagery [6]. Several studies [8, 9, 13] have been conducted on ”color+depth” representation of 3D video. However, in [8, 9] the employed 2D metrics do not consider 3D features such as depth and characteristics of stereo vision.

Da Silva et al. [7] propose a full reference metric which measures quality of symmetric and asymmetric coded stereoscopic video. Two main types of artifacts are considered in this solution: structural distortions and blurring. It achieves correlation coefficients of nearly 0,95. 16 non-expert observers participated in the subjective experiment connected with this research.

Seo *et al.* [11] propose a full-reference metric based on an algorithm which calculates the scores of evaluation items (blocking artifacts, blurring in edge regions and inter-view similarity), and performs a weighted summation for those scores. The authors consider disparity of information (the difference between left and right images) as a 3D factor for the assessment of compressed stereoscopic video. They compare their results with conventional quality metrics, and their metric has a higher correlation than others (approx. 0.95). Twenty people who were able to perceive 3D video took part in the subjective experiment.

Xing *et al.* [13] propose a 3D metric which has a Pearson correlation of 0.88 and is based on predicting quality levels of crosstalk perception in stereoscopic images. The basis for building this metric was an understanding of three main factors: crosstalk level, camera baseline, and scene content. Crosstalk is an error in separating the left and right channel, making the imagery unpleasant and annoying for the end-user. In the subjective experiment, 28 subjects took part.

Zhu and Wang [14] propose a 3D metric based on state-of-the-art physiological and psychological understanding of the human visual system (HVS). They analyze several main properties - such as Contrast Sensitivity Function, Masking effect, Multi-channel Mechanism and Depth Perception - and propose a multi-channel vision model based on 3D wavelet decomposition. The resulting metric has a correlation of approx. 0.65. The number of subjects who took part in the test is unknown.

Currently available research reports on QoE metrics for 3D video do not include efforts in building bitstream metrics, available to use in real-time, without need for full decoding.

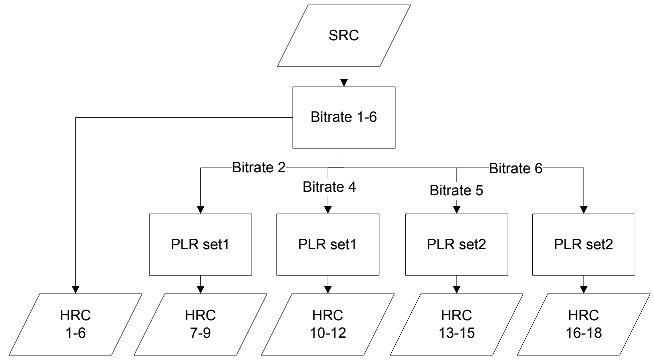

Fig. 1. Generation of HRCs.

This kind of solution is the most interesting for telecoms operators.

3. TEST MATERIALS

In this section, we provide a short description of the experiment. The results of this experiment were used as the basis for building a new quality metric. A detailed description of how the test sequence was prepared can be found in [10].

The raw sequences used are short 3D videos of 4K in size, shared by the NTT Corporation of Japan thanks to a collaboration with the Video Quality Experts Group (VQEG). They were divided into scenes, and nine of them were selected for conversion to FullHD in the ”up side down side” manner. As a result, each sequence had a size of 1920x2160 pixels.

3.1. Source Video Sequences (SRC)

Source sequences are generated with special modifications imposed by Hypothetical Reference Circuits (HRC) shown in figure 1.

All of the SRCs are coded with six different bitrates: HRC 1-6 (1 Mbit/s, 2 Mbit/s, 3 Mbit/s, 4 Mbit/s, 5 Mbit/s and 50 Mbit/s). Additionally, three levels of packet loss are set and applied to four of six bitrate values (HRC 7-18). By this modification, 162 PVSes are used during the experiment. Examples of frame sequences are presented in Figure 2. Scenes were selected in order to show a diversity of camera movement, depth effect intensity and motion level.

4. EXPERIMENT OVERVIEW

Subjects were selected by a recruitment agency which prechecked their stereo vision acuity, using the Randot Stereo Test. A positive result was a necessary condition for taking part in the experiment. All of the 30 subjects were naive observers, meaning that their profession was not connected with television picture quality or assessing services. Each subject

Fig. 2. Frames from source sequences — first row: digest; second row: okuagi; third row: okunai.

had either normal vision acuity or was supported with proper corrective glasses or contact lenses.
Each subject individually followed the ACR (Absolute Category Rating) scenario described in Recommendation ITU-T P.910 [1]. After watching a ten-second video sequence, the subjects had fourteen seconds in which to assign a score on the MOS scale, before the next sequence was presented on the display.

Viewing conditions of the test room followed Recommendation ITU-R BT.500-12 [2] and Recommendation ITU-T P.910 [1]. For the subjective experiment, a 24" BENQ 3D display with shutter glasses was used to display and view 3D video content. The test room is presented in Figure 3 .

Fig. 3. Test environment

5. PROPOSED METRIC

The basis for the metric was a linear model. Parameters extracted from the bitstream were the most important part of it. In Figure 4, a concept diagram of this metric is presented. The QoE measurement module analyzed all packets received by a client, then it detected and estimated packet loss ratio, and estimated bitrate and content dynamism. According to [10], the subjective score is strongly sensitive and depends on the type of transmitted content. The content dynamism could be estimated relatively easy by comparing the size of P and I slices used in the coding process.

The component extraction algorithm is presented in Figure 5. Headers were analyzed on the basis of received packet sequences. The number of lost packets was derived from RTP header analysis - rather than analysis of H264 stream headers, which provides a distinction between I and P slices and determines their sizes. This led to the estimation of the *bitrate* value.

Due to the fact that the scope of bitrates is relatively high,

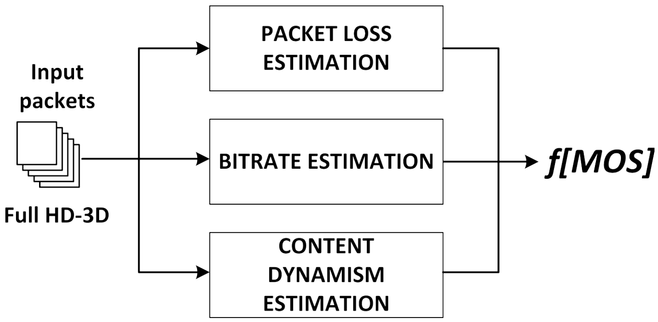

Fig. 4. Components of QoE metric for 3D video

this metric is built specially for bitrates lower than 4 Mbit/s and bitrates higher than 4 Mbit/s.

$$f_{low}[MOS] = a_{low}PLR + b_{low}bitrate + c_{low} \quad (1)$$

$$f_{high}[MOS] = a_{high}PLR + b_{high}Psize + c_{high} \quad (2)$$

where

$$PLR = \frac{(number\ of\ lost\ packets) * 100}{number\ of\ packets}$$

$$bitrate = 8 * (29 * Psize + Isize)$$

$$Psize = average\ size\ of\ slice\ P$$

$$Isize = average\ size\ of\ slice\ I$$

$$a, b, c = defined\ coefficients$$

Coefficients are determined by the results of the subjective experiment.

Figure 6 presents predicted MOS values as a function of observed MOS values for bitrates lower than 4 Mbit/s. The proposed metric for this lower scope of bitrates has a correlation of 0.71. In Figure 7, we present the predicted MOS values as a function of observed MOS values for bitrates higher than 4 Mbit/s. The proposed metric for this higher scope of bitrates has a correlation of 0.89. Unification of this model for the whole scope of bitrates causes a decline in precision.

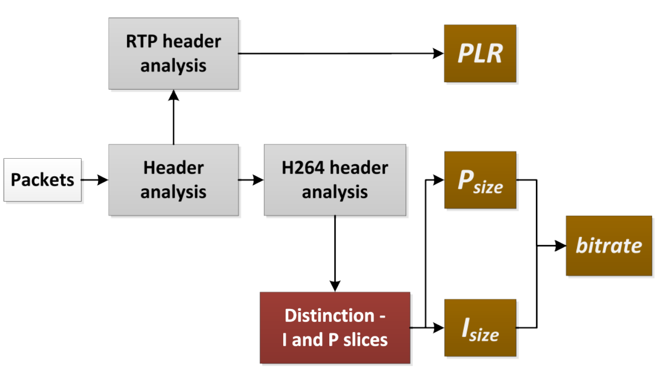

Fig. 5. Design of the module for measuring QoE for 3D video

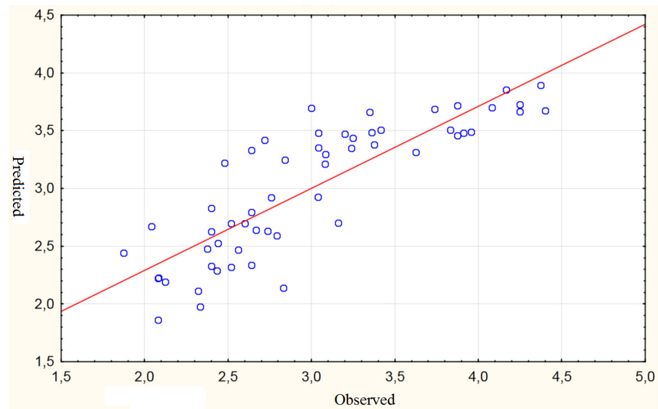

Fig. 6. Predicted MOS values as a function of observed MOS values for bitrates lower than 4 Mbit/s.

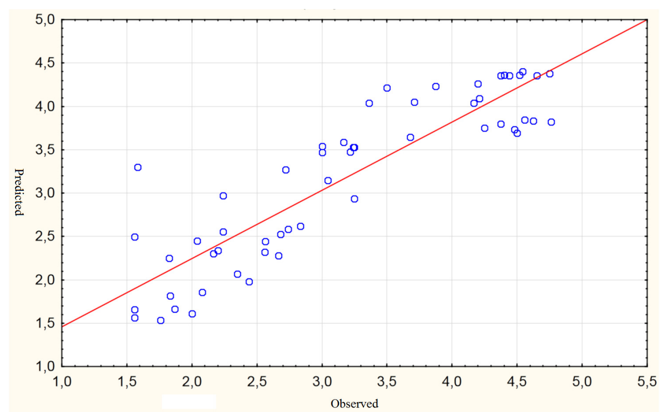

Fig. 7. Predicted MOS values as a function of observed MOS values for bitrates higher than 4 Mbit/s.

6. CONCLUSION

The design of QoE metrics for 3D video is a very difficult task, due to a large variety of factors influencing the perception of the 3D effect. In this paper, we have presented a metric built on parameters extracted from a bitstream. Therefore, this solution can be classified as a 'bitstream no-reference metric'. To the best of our knowledge, this metric is original, because it is the first to predict QoE for HD-3D in 'side by side' representation. From a practical point of view, its main advantage is that the bitstream does not have to be fully decoded, making real-time quality control possible.

The results of this experiment are promising, therefore further subjective evaluations need to be conducted in order to improve the metric's precision, especially in terms of content influence.

7. REFERENCES

[1] Itu-t p.910, subjective video quality assessment methods for multimedia applications, 1999.

[2] Itu-r bt.500-12, methodology for the subjective assessment of the quality of television pictures, 2009.

[3] 3d display technology and market forecast report. Technical report, NPD DisplaySearch, 2010. January.

[4] Cisco visual networking index: Forecast and methodology, 2011 - 2016. Technical report, Cisco Systems, Inc., May 2012.

[5] Alexandre Benoit, Patrick Le Callet, Patrizio Campisi, and Romain Cousseau. Quality assessment of stereoscopic images. *EURASIP Journal on Image and Video Processing*, 2008(1):659024, 2008.

[6] P. Campisi, Callet P.L., and Marini E. Stereoscopic images quality assessment. Poznan, Poland, September 2007.

[7] V. De Silva, H.K. Arachchi, E. Ekmekcioglu, and A. Kondoz. Toward an impairment metric for stereoscopic video: A full-reference video quality metric to assess compressed stereoscopic video. *Image Processing, IEEE Transactions on*, 22(9):3392–3404, 2013.

[8] C. T E R Hewage, S.T. Worrall, S. Dogan, S. Villette, and A.M. Kondoz. Quality evaluation of color plus depth map-based stereoscopic video. *Selected Topics in Signal Processing, IEEE Journal of*, 3(2):304–318, 2009.

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

[10] D. Juszka, L. Janowski, and Z. Papir. Ocena jakoci wideo 3d-hd znieksztaconego przez kompresj i straty pakietw. *Przegld Telekomunikacyjny, Wiadomoci Telekomunikacyjne*, 85(4):294297, 2012.

[11] Jungdong Seo, Xingang Liu, Donghyun Kim, and Kwanghoon Sohn. An objective video quality metric for compressed stereoscopic video. *Circuits, Systems, and Signal Processing*, 31(3):1089–1107, 2012.

[12] Wa James Tam, Lew B. Stelmach, and Philip J. Corriveau. Psychovisual aspects of viewing stereoscopic video sequences. volume 3295, pages 226–235, 1998.

[13] Liyuan Xing, Junyong You, T. Ebrahimi, and A. Perkis. A perceptual quality metric for stereoscopic crosstalk perception. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 4033–4036, 2010.

[14] Zhongjie Zhu and Yuer Wang. Perceptual distortion metric for stereo video quality evaluation. *WSEAS Trans. Sig. Proc.*, 5(7):241–250, July 2009.