Extended No Reference Objective Quality Metric for Stereoscopic 3D Video

Yi Han[1,3], Zhenhui Yuan[2], and Gabriel-Miro Muntean[3]
[1] School of Computer Science, University College Dublin, Ireland
[2] Huawei Technologies Co., Ltd
[3] School of Electronic Engineering, Dublin City University, Ireland
yi.han@ucdconnect.ie, yuanzhenhui@huawei.com and gabriel.muntean@dcu.ie

***Abstract*— Currently, three-dimensional (3D) video is gaining increasing popularity by providing immersive user experience. Compared against conventional 2D video, 3D video excels at bringing a "live" scene closer to the users, and/or trying to place the users in the environment of the displayed content. However, streaming 3D video sequences over the IP networks is challenging due to the impact of dynamic network conditions on user quality. Accurate objective 3D video quality assessment is critical for advanced real-time video streaming adaptation solutions. Most state-of-the-art objective 3D video quality metrics are reference-based and require access to the original 3D video sequences, which is not possible for the real-time applications. This paper proposes the extended No-reference objective 3D Video Quality Metric (eNVQM) for real time 3D video quality assessment. eNVQM establishes a correlation between network packet loss and stereoscopic 3D video quality and was tuned according to extensive subjective testing results. Performance of eNVQM is studied in comparison with two state-of-the-art objective video quality metrics: structural similarity index (SSIM) and video quality metric (VQM).**

***Index Terms*—3D video; objective quality assessment; non-reference; stereoscopic**

I. Introduction

APART from the classic 2D video content, the 3D video also supports dissemination of the sense of depth, significantly enhancing the user viewing experience. Due to the advanced development of image processing, display technologies, and digital video coding approaches (e.g. H.264/AVC, H.264/SVC and Multiview Video Coding (MVC)), 3D video techniques have been widely deployed in various application areas, including 3D movies, 3DTV, 3D gaming, etc. On the other hand, the increasing capacity and speed of both core and access networks support the delivery of highly popular 3D video to a large user base, including mobile, and opening new opportunities for diverse applications beyond the traditional theatre-based 3D movie shows, such as mobile 3D video streaming, 3D video chat, and 3D conferencing.

The 3D video has brought a revolutionary enhanced viewing experience closer to the video users; however in addition to the challenges that already exist in relation to 2D video, there are challenges specific to 3D video for providing users good quality levels. This is also because the depth sense associated with 3D video may enhance or decrease the overall 3D viewing experience depending on the effect of the image compression and delivery. The 3D video content consists of separate video frames for the left and right eyes, which form separate left and right video streams/layers, respectively, which often have redundant information that is reduced during compression by various algorithms. Thus diverse network impairments that affect either layer (left or right) of the 3D video content may result in different levels of degradation of the 3D video quality. Additionally, the impact of encoding at different bitrates and frame rates on the 3D video varies from that on the 2D video.

Network delivery of 3D video at good quality level is challenging mostly due to the network dynamic conditions. Despite the development of various network solutions, often the performance of the video delivery is affected, especially for mobile and interactive applications. 2D video delivery research showed that there is a need for more advanced solutions, including adaptive delivery schemes [21][27][28], which are aware of the network conditions and adjust dynamically the video delivery process in order to maintain good user perceived quality levels. These solutions require accurate real-time estimation of the user perceived Quality of Experience (QoE).

Some research has been conducted to estimate or measure the QoE levels of the 3D video. Subjective methods involving people evaluating the 3D video quality provide highly accurate results as they directly reflect human quality perception levels. However, these methods require carefully controlled environments, cannot be done online during transmission, require important human resources and are time consuming. Objective quality assessment methods can be performed in real-time and are therefore preferred, but are less accurate. Recently, several objective quality metrics for 3D video have been proposed [1] – [6], but they lack accuracy, mainly due to the fact that the human visual system (HVS) is difficult to model in pixels and depth, and is also affected by other factors such as human eye comfort level, viewing distance, etc. Furthermore, the existing objective 3D video quality assessment methods are highly dependent on the video content and do not consider network impairments. The quality metrics widely used for 3D video quality assessment employ 2D video quality metrics, including PSNR [7], SSIM [8], and VQM [9]. These 2D video quality metrics can be used to measure the video quality for the left and right views separately and the 3D video quality can be derived by considering different weights for the two views [2][11]. However, the above approaches require the usage of both decoded and original video sequence in order to analyse the blockiness, blurring, and depth information in the decoded video content [12]. Such quality assessment can only be done off-line when the transmission is over, and thus is not suitable for real-time assessment. The no reference PSNR [10] can be used online, but it is based on assumptions related to the manner bitrate and loss affects encoded 2D video and therefore its accuracy in a 3D context will be very limited.

This paper investigates the impact of variable network conditions on the perceived quality of 3D video content. The proposed **extended No-reference 3D Video Quality Metric (eNVQM)** extends the previously described NVQM [14] by considering not only network-related parameters (e.g. loss), but also the video characteristics such as bitrate and frame rate [13]. eNVQM employs the principle of ITU-T G.1070 model proposed for 2D video quality assessment [15], butis designed for 3D video quality estimation. eNVQM is derived based on real subjective testing results and can be used for proactive adaptive 3D video transmissions.

The rest of this paper is organized as follows. Section II presents the current subjective and objective 3D video quality metrics and Section III introduces the eNVQM model. Section IV describes the experimental setup and the experimental results are analysed in Section V. In the end, Section VI concludes the paper.

II. Related Work & Technical Background

Measuring 3D video quality has already been investigated using 2D objective video quality metrics such as PSNR, SSIM and VQM. Authors of [1] have shown that, by averaging the separate results of left and right views, VQM can predict the overall image quality, while PSNR and SSIM results have better correlation with the 3D video depth perception than those of VQM. Study of the quality evaluation of colour plus depth map-based 3D video using these 2D video quality metrics is described in [2], in which the 3D video quality is an average score of the rendered left and right video using an innovative Depth-Image-Based Rendering (DIBR) technique. Rather than using the same weight for left and right views, the authors of [4] assign weights of 1/3 and 2/3 of the PSNR score to the left and right views respectively.

A new perceptual quality metric (PQM) was proposed in [6]. Being more sensitive to image degradation and error quantification that happen at pixel level than at sequence level, PQM shows better results for 3D video quality in comparison with VQM. The impact of eye dominance on the perceived 3D video quality is modelled in [3], which is based on spatial frequency by chopping the images into small 4*4 blocks. The edge distortion in depth and colour 3D videos also has a significant impact on the 3D video quality and this has been modelled in the colour and sharpness of edge distortion measure (CSED) proposed in [5]. An objective model in [25] predicts the quality of lost frames in 3D video streams based only on the estimated lost frame size. The authors in [26] also evaluated stereoscopic 3D video quality using 2D objective metrics, including PSNR, SSIM and VIFP. Their results show that the colour perception is a dominant in the overall 3D video quality while the depth has less impact.

These 3D video quality metrics have different accuracy levels and advantages. However they all require full reference of the original video source and differ from our proposed non

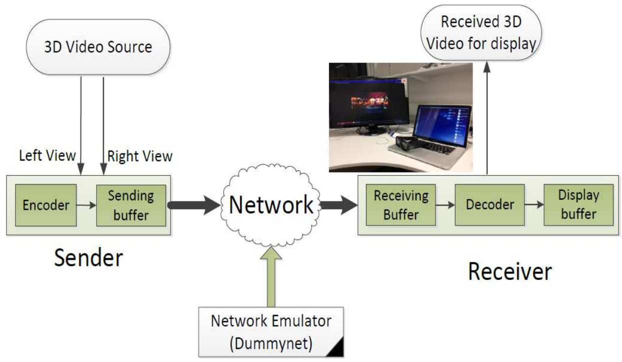

Figure 1. Experiment Framework

reference network-based metric, which does not require the presence of either the original or degraded 3D video.

Stereoscopic 3D video contains left and right view video for left and right eyes, respectively. The two views can either be stored in a single video file or two separated video files. These two views are synchronized and played by a 3D supported player simultaneously, providing human viewer two perspectives of the same scene with a minor deviation. This deviation gives the perception of 3D depth while it is processed by human brain. When stereoscopic 3D video is transmitted over the network, the two views are combined into a frame sequential stream. The video frames are stacked one by one from left and right views in a frame sequential manner [17].

III. Proposed extended 3D Video Quality Model

The idea behind eNVQM is to investigate properly the relationship between network packet loss, and the 3D video bitrate and frame rate and model it. The model has three input variables, and thus it requires accurate mapping at each of these three dimensions of the model. ITU-T G.1070 [15] has defined a similar model for 2D video. eNVQM employs the idea and major coefficients of the ITU-T G.1070 model, and proposes a new model for 3D video considering both colour and depth information.

A. ITU-T G.1070 2D Video Quality Metric

The ITU-T has standardized a user opinion model for 2D video-telephony applications in G.1070. It estimates the 2D video quality in telephony applications by considering the network impairment parameters (i.e. packet loss in video) and encoding parameters, including codec type, video format, key frame interval, and video display size.

The 2D video quality is modeled by equation (1):

$$V_q = 1 + I_{coding}\, e^{-\frac{Ppl_V}{D_{PplV}}} \qquad (1)$$

where Ppl_V represents packet loss rate, D_{PplV} expresses the degree of video quality robustness due to packet loss, and I_{coding} calculates the basic video quality affected the coding impairment that is influenced by video bitrate (Br_V is expressed in kbps) and video frame rate (Fr_V is measured in fps). Note (1+ I_{coding}) represents the video quality when the packet loss is 0%.

In the G.1070 model there are twelve coefficients which are derived from subjective 2D video tests and are dependent on the video coding, and display size. The methodology for deriving the coefficients in the model is given in [15]. The recommendation gives five sets of coefficients for different display sizes for MPEG-4 and ITU-T H.264, respectively. In the standard, the related accuracy of the predicted video quality was evaluated by the Pearson product-moment correlation [22].

The derivation of the proposed eNVQM for 3D is shown in the next sub section.

B. Extended No Reference 3D Video Quality Metric (eNVQM)

The proposed eNVQM is designed to estimate 3D video quality based on packet loss rate, 3D video bitrate and 3D video frame rate in a no reference manner.

The stereoscopic 3D video consist of left and right views, each similar to a 2D video. The left and right views in stereoscopic 3D video are directed to left and right eyes of the human observers with various display technologies. The differences in the two views produce illusion in human perception and this provides the observer the sense of depth in a 3D space. During network transmission of a 3D content, any information loss affects the video quality. If information is lost in either left or right view for the same video frame, it might be compensated from the other view, reducing the quality loss. When the information cannot be compensated from the other view, it may affect the display of the other view, resulting in an impaired 3D displayed frame and thus decreasing the 3D video quality. These are the reasons for which we believe that the network impairment has different impact on 3D video than that of 2D video.

The 2D video quality metric described in G.1070 provides a good methodology of mapping bitrate, frame rate and packet loss to the 2D video quality MOS. Based on this, we propose for eNVQM the formula from equation (2), where I^{3D}_{coding} is composed of two additive natural logarithms for both frame rate and bitrate, reflecting the video quality when packet loss (Ppl_V) is 0%. The remaining part of eNVQM formula represents the effect of packet loss on the video quality when considering 3D video frame rate and bitrate. The equations are presented below:

$$V^{3D}_{q} = 1 + I^{3D}_{coding}\, e^{-\frac{Ppl_V}{D^{3D}_{PplV}}} \quad (2)$$

$$I^{3D}_{coding} = a_1 \ln(Fr_V) + a_2 \ln(a_3 + a_4 Br_V) \quad (3)$$

$$D^{3D}_{PplV} = a_5 + a_6 * e^{-\frac{Fr_V}{a_7}} + a_8 * e^{-\frac{Br_V}{a_9}} \quad (4)$$

Equations (2)-(4) are used to estimate the quality of both colour and depth components of the 3D video: V^{3D}_{color} and V^{3D}_{depth}. Two sets of coefficients A = {a_1, a_2,…, a_{12}} are derived from subjective 3D video tests of colour (A_{color}) and depth perception (A_{depth}), respectively. a_1 and a_2 reflect the effect of frame rate and bitrate respectively when there is no packet loss. a_3 and a_4 quantifies the contribution of bitrate so that both frame rate and bitrate can be represented in a balanced manner in the overall formula. There is no need for frame rate to have similar coefficients to bitrate because the frame rate (10 ~ 60 fps) scales faster than bitrate (1~10 Mbps) and thus it is well enough represented by the natural logarithm. The coefficients a_5 to a_9 are used to map different scales of frame rate and bitrate on the scale of packet loss rate, respectively. Coefficients $a_{1,}$ $a_{2,\,\ldots,}$ and a_9 are dependent on the codec type, video format, and display size.

Furthermore, the overall 3D video quality needs to combine colour and depth qualities. It is assumed that there is an additive effect of depth perception on the colour perception in terms of the 3D video quality, so equation (5) is employed:

$$V^{3D}_{q} = xV^{3D}_{color} + yV^{3D}_{depth}, \qquad x + y = 1 \quad (5)$$

where x and y give different weights to colour and depth perception, respectively. The values of x and y are derived from three other perceptual factors considered in the subjective testing, reflecting eye comfort level, whether the 3D video is enjoyable, whether the 3D effect enhances the experience.

IV. Experimental Setup

An extensive set of experiments are designed to study the relationship between the network characteristics (i.e. packet loss), 3D video encoding settings (i.e. framerate, bitrate) and the perceived 3D video quality, independent from the video content. Different network scenarios with varying network packet loss ratios were generated. The independence of the video content is ensured by making use of high number of 3D video samples with different contents, each encoded with a set of encoding settings.

Five 3D video clips with content belonging to different scenarios with diverse motion complexity levels are selected from the database in [23], as shown in Table I. The duration of the selected video clips varies from 6 to 14 seconds, according to [16]. These video clips are MPEG-4 SVC encoded with high (4 Mbps), medium (3 Mbps), and low (2 Mbps) average bitrates following the IPPP sequence format and have frame rates of 18 fps and 11 fps, targeting mobile applications. The clip scene scenarios include running, driving, swimming, etc. as indicated in Table I.

Figure 1 shows the test topology of the experiment. Each 3D video clip is encoded and transmitted using RTP over the network separately from *Sender* to *Receiver* using the VLC media player. *Dummynet* is used to control the desired packet loss in the network. The *Receiver* receives the stream sent over the impaired network, and decodes the stream to video files for left and right video, creating a *video clip pair*, separately.

TABLE I
VIDEO SAMPLES

Clip No.	Motion complexity level	Scene scenarios	Duration (seconds)	Scene
1	High	Running	9	
2	High	Driving	14	
3	Medium	Swimming	13	
4	Medium	Dancing	6	
5	Low	Kissing	8	

Wireshark is used at the receiver side in order to capture the transmitted stream and calculate the packet loss. 11 network loss scenarios are created: 0%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 8% and 10%. More scenarios were considered and studied in lower packet loss range (less than 5%) to allow for better accuracy. Overall there are 11 * 3 * 2 * 5 = 330 video clip left-right pairs transmitted using the experiment.

Subjective testing is conducted with 40 volunteers. 330 videos are divided into 10 groups, each containing 33 videos randomly selected from different video contents, packet loss, bitrate and framerate. Thus each individual clip has at least 4 results from 4 different observers. The clips are displayed on a 27 inches *3D Asus VG278* monitor with resolution 1920x1080 pixels, and the *3D vision*[2] support enabled from *Nvidia*. The participants are required to wear a pair of 3D vision[2] wireless active shutter glasses. As suggested by the monitor manufacturer, the viewing distance is set to 1 m. The tests are conducted in a 5m x 5m quiet room, having the monitor away from window to avoid additional unnecessary light for optimum viewing experience. Each participant is asked to grade their overall 3D experience, 3D depth experience, eye comfort, 3D enjoyable level, 3D effect enhancement level (whether 3D effect enhance the overall viewing experience). The grading uses the 1 (bad) to 5 (excellent) MOS scale.

V. RESULT ANALYSIS

The results collected from the subjective tests consist of grading marks for the 330 video clips, each having a particular combination of bitrate, frame rate, video content, and packet loss rate.

The goal is to derive a mapping from packet loss, bitrate, and frame rate values to an estimation of user perceptual 3D video quality.

75% of the subjective results are used for the model derivation and 25% of the test results are reserved to allow for the verification of the derived model. A fitting curve is derived as shown in Figure 2 and Figure 3 for colour and depth

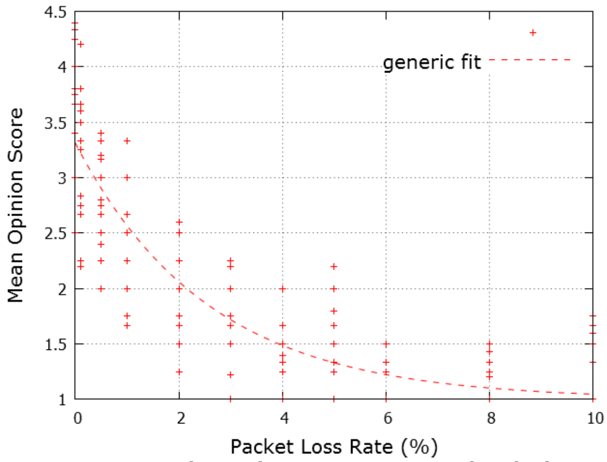

Figure 2. Mapping colour quality perception vs. network packet loss rate

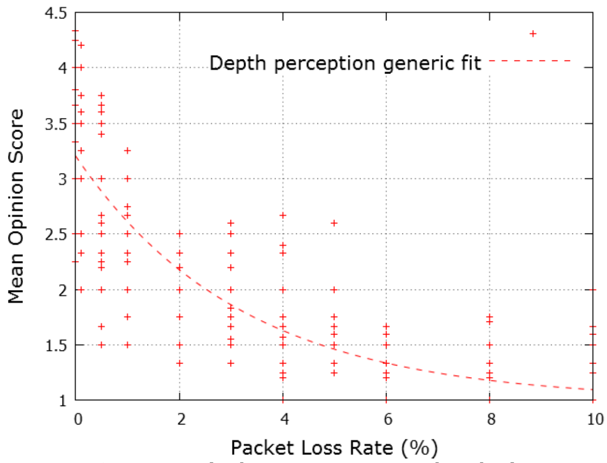

Figure 3. Mapping depth perception vs. network packet loss rate

parameters, respectively. The coefficients a_1, to a_5 are calculated for A_{color} and A_{depth} , following by the steps described in ITU-T G.1070, respectively. The method involves calculating some of coefficients by having only one of them variable and keeping the other ones fixed. The coefficients are computed using the Least Square Approximation (LSA) [24].

The raw subjective results are processed in order to eliminate outliers. For each clip, an outlier result is considered if the score is more than 2 grades adrift from the median MOS of all the values recorded from all this clip's viewers. However, when considering packet loss scenarios, for each case there are 22 clips (75% of 30 clips) with different content, bitrates and frame rates (with packet loss rate fixed). Among these 22 results, the highest and lowest 10% of them are considered outliers and are removed. The same process is performed for both overall colour and depth perception, respectively. The mappings between packet loss and the two types of perception are shown in Figure 2 and Figure 3, respectively. The corresponding coefficients for colour and depth models instantiated from equations (2)-(4) are listed in Table II.

In order to verify the correctness of the model, the remaining 25% of the subjective results are used to compute Pearson correlation with the eNVQM results. The model uses inputs

TABLE II
COEFFICIENTS COMPUTED FOR eNVQM

	colour	*depth*
a_1	0.09136	0.08751
a_2	1.11132	1.05853
a_3	0.93128	0.93067
a_4	1.79391	1.7921
a_5	-1.24607	-0.46754
a_6	0.01436	1.67570
a_7	33.775	33.03
a_8	2.17023	0.39725
a_9	5.37876	4.45855

TABLE III
VERIFICATION OF eNVQM COEFFICIENTS - PEARSON CORRELATION OF eNVQM AND SUBJECTIVE RESULTS

	25% subjective results		*75% subjective results*		*100% subjective results*	
	colour	depth	colour	depth	colour	depth
Correlation	0.873	0.785	0.916	0.903	0.942	0.935

with the same frame rate, bitrate and packet loss rate as in the clip presented to the observer. The comparison is shown in Table III. The slightly lower correlation when 25% of the results are considered is caused by the low number of results available and higher variations. The relative high level of correlation indicates that our derived model coefficients are valid and reliable.

Next, weights for colour and depth for 3D video are determined by making use of three additional factors: eye comfort, 3D enjoyment level, and 3D effect enhancement level. The same process of removing outliers in each clip is followed, but outliers when considering a particular packet loss rate are retained, as no fitting curve is required in this step. Giving different weights to colour and depth, the overall scores are compared against the results of the above three factors. Each result set organizes data for a particular packet loss rate in each row with a combination of bitrate and frame values in each column. Correlations are computed for each column pairs containing subjective results and grading marks for the above factors. Finally the average correlations over all packet loss rates are calculated. This is done for each of the three subjective factors considered. The highest average correlation of these factors is considered to determine the weights of x, y for colour and depth perception. The trend follows a 2[nd] order polynomial function, in which y is replaced by (1-x):

$$Correlation = -0.0026x^2 + 0.0046x + 0.8644 \qquad (6)$$

The function of the correlation trend is a parabola of x (since y = (1-x) and its vertex is at x=0.885, giving the highest correlation of 0.866434615. Thus equation (6) can be expressed as in equation (7):

$$V_{3D_q} = 0.885 * V_{colour} + 0.115 * V_{depth} \qquad (7)$$

where V_{colour} and V_{depth} are calculated using equations (2) - (4) and the coefficients in Table II.

The eNVQM model takes three input variables: frame rate, bitrate and packet loss rate. The output of eNVQM is expressed in terms of MOS and refers to the human perception of 3D video quality. Fig. 4 illustrates eNVQM variation against bitrate and framerate when the packet loss is 0%, 1% and 3%. It can be noted how MOS increases as bitrate and frame rate become larger and how the effect of bitrate growth is larger in terms of MOS than a frame rate increase. And the effect of bitrate and frame rate differs for different packet loss rates. Fig. 5 shows specifically eNVQM variation against loss rate and frame rate at a fixed bitrate of 4 Mbps. It is also interesting to see that for lower range frame rates, MOS drops more rapidly relative to packet loss growth, while MOS drops smoothly for higher range frame rates.

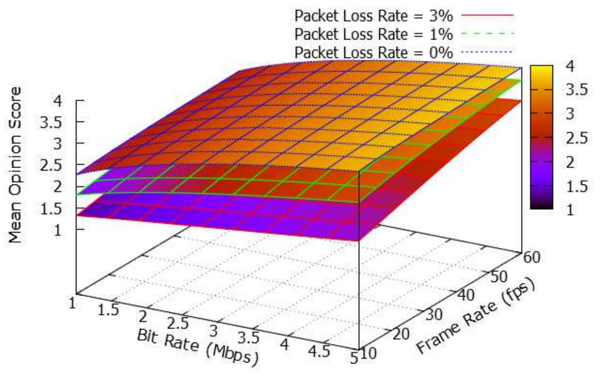

Figure 4. eNVQM 3D video Quality with 0% packet loss

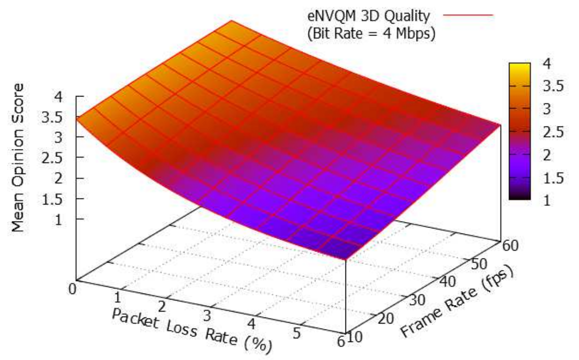

Figure 5. eNVQM 3D video Quality with bitrate 4 Mbps

Other research works from the literature employ SSIM and VQM for objective 3D video quality assessment. Despite our reluctance regarding the use of 2D metrics to assess 3D video quality, in order to compare the performance of the proposed eNVQM to other models, SSIM and VQM results applied to the 3D video are shown next. The same weights for both left and right views [1] [2] were applied in order to compute the objective 3D video quality. MSU VQMT [18] was used as computational tool. Since SSIM and VQM use different scales

TABLE IV
PEARSON CORRELATION

	SSIM	VQM	eNVQM
Correlation	0.911	0.932	0.944

from MOS, normalization methods described in [19] and [20] were employed, respectively. The original and degraded sample pairs were compared by VQMT for the left and right views, and the average scores of both views converted to MOS scale were compared with the results of eNVQM. Considering 100% of the subjective results, the Pearson correlations with the subjective test results are listed in Table IV. This correlations show that by using eNVQM higher accuracy in predicting the perceived 3D video quality can be obtained in comparison when using other reference methods.

VI. CONCLUSION

This paper proposes the extended no reference objective video quality metric (eNVQM) for the online assessment of stereoscopic 3D video quality. eNVQM estimates the 3D video quality using frame rate, bitrate and network packet loss rate. Pearson correlation shows that eNVQM has better accuracy in terms of human perception in 3D video, comparing against two current common assessment methods SSIM and VQM. eNVQM is perfect for adaptive 3D video transmissions as it can quickly estimate the current video quality so that delivery adjustment actions can be taken at the earliest possible point, increasing user perceived quality levels.

ACKNOWLEDGMENT

This research was supported in part by the Telecommunications Graduate Initiative (TGI), funded by the Higher Education Authority under the Programme for Research in Third-Level Institutions (PRTLI) Cycle 5 and co-funded under the European Regional Development Fund (ERDF).

REFERENCES

[1] Yasakethu, S. L. P., Hewage, C. T., Fernando, W. A. C., and Kondoz, A. M., “Quality analysis for 3D video using 2D video quality models,” *IEEE Trans. Consum. Electron.*, vol. 54, no. 4, pp. 1969-1976, Nov. 2008.

[2] Hewage, C. T., Worrall, S. T., Dogan, S., Villette, S., and Kondoz, A. M., “Quality evaluation of colour plus depth mapbased stereoscopic video,” *IEEE. J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp 304-318, Apr. 2009.

[3] Lu, F., Wang, H., Ji, X., and Er, G., “Quality assessment of 3D asymmetric view coding using spatial frequency dominance model,” in *Proc. 3DTV Conference*, Potsdam, Germany. pp. 1-4, May 2009.

[4] Ozbek, N., and Tekalp, A. M., “Unequal inter-view rate allocation using scalable stereo video coding and an objective stereo video quality measure,” in *Proc. IEEE Int. Conf. Multimedia and Expo*, pp. 1113-1116, Monterrey, México, Apr. 2008.

[5] Shao, H., Cao, X., and Er, G., “Objective quality of depth image based rendering in 3DTV system,” in *Proc. 3DTV Conference*, May 2009.

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

[7] Huynh-Thu, Q., Ghanbari, M., “Scope of validity of PSNR in image/video quality assessment,” *Electronics Letters* , vol.44, no.13, pp.800-801, Jun. 2008.

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

[9] Pinson, M. H., and Wolf, S., “A new standardized method for objectively measuring video quality,” *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp.312 -322, Sept. 2004.

[10] Lee, S.-B., Muntean, G.-M., Smeaton, A.F., “Performance-aware replication of distributed pre-recorded IPTV content,” *IEEE Transactions on Broadcasting*, vol.55, no.2, pp.516-526, Jun. 2009.

[11] Yasakethu, S. L. P., Hewage, C. T., Fernando, W. A. C., and Kondoz, A. M., “Quality analysis for 3D video using 2D video quality models,” *IEEE Trans. Consum. Electron.*, vol. 54, no. 4, pp. 1969-1976, Nov. 2008.

[12] Moller, S., Wai-Yip Chan, Cote, N., Falk, T. H., Raake, A., Waltermann, M., “Speech Quality Estimation: Models and Trends,” *IEEE Signal Processing Magazine*, vol.28, no.6, pp.18-28, Nov. 2011.

[13] Yasakethu, S. L. P., Hewage, C. T., Fernando, W. A. C., and Kondoz, A. M., “Quality analysis for 3D video using 2D video quality models,” *IEEE Trans. Consum. Electron.*, vol. 54, no. 4, pp. 1969-1976, Nov. 2008.

[14] Han, Y., Yuan, Z., Muntean, G.-M., "No reference objective quality metric for stereoscopic 3D video," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp.1-6, Jun. 2014.

[15] ITU-T G.1070, “Opinion model for video-telephony applications”, Jul. 2007.

[16] ITU-T P.913, “Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment”, Apr. 2014.

[17] Lang, M., Hornung, A., Wang, O., Poulakos, S., Smolic, A., and Gross, M., “Nonlinear disparity mapping for stereoscopic 3D,” *ACM Transactions on Graphics (TOG)*, vol. 29, issue 4, no.75, Jul. 2010.

[18] MSU Video Quality Measurement Tool. Available : http://compression.ru/video/quality_measure/video_measurement_tool_en.htm.

[19] Zinner, T., Abboud, O., Hohlfeld, O., Hossfeld, T., and Tran-Gia, P. (2010, March). “Towards qoe management for scalable video streaming,” *21th ITC Specialist Seminar on Multimedia Applications-Traffic, Performance and QoE* , pp. 64-69, Mar. 2010.

[20] Dymarski, P., Kula, S., and Huy, T. N., “QoS Conditions for VoIP and VoD,” *Journal of Telecommunications & Information Technology*, vol. 2011, no.3, pp.29-37, Mar. 2011.

[21] Gopalakrishnan, V.; Bhattacharjee, B.; Ramakrishnan, K.K.; Jana, R.; Srivastava, D., "CPM: Adaptive Video-on-Demand with Cooperative Peer Assists and Multicast," *IEEE INFOCOM*, , pp.91-99, Apr. 2009.

[22] Adler, J., Parmryd, I. “Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient”. *Cytometry Part A*, 77(8), 733-742. 2010.

[23] Hadfield, S., Bowden, R., “Hollywood 3D: Recognizing Actions in 3D Natural Scenes,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.3398-3405, Jun. 2013.

[24] Sarkar, B., Menq, C. H., “Smooth-surface approximation and reverse engineering”, *Computer-Aided Design*, 23(9), 623-628, 1991.

[25] Feitor, B., Assuncao, P., Soares, J., Cruz, L., Marinheiro, R., "Objective quality prediction model for lost frames in 3D video over TS," *IEEE International Conference on Communications Workshops (ICC)*, pp.622-625, Jun. 2013.

[26] Wang, K., et al.. “Stereoscopic 3D video coding quality evaluation with 2D objective metrics”. *In Proc. SPIE Electronic Imaging* (Vol. 8648), Mar. 2013.

[27] Yuan, Z. and Muntean, G.-M., "A Prioritized Adaptive Scheme for Multimedia Services over IEEE 802.11 WLANs", *IEEE Transactions on Network and Service Management*, no.4, vol.10, pp.340 - 355, Dec. 2013.

[28] Yuan, Z., Venkataraman, H. and Muntean, G.-M., "iPAS: An User Perceived Quality-based Intelligent Prioritized Adaptive Scheme for IPTV in Wireless Home Networks," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Shanghai, China, , pp.1-6, Mar. 2010.