REDUCED-REFERENCE QUALITY METRIC FOR 3D DEPTH MAP TRANSMISSION

Chaminda T.E.R. Hewage and Maria G. Martini

Wireless, Multimedia and Networking Research Group, Faculty of CISM, University of Kingston, Kingston upon Thames, Surrey, UK.

ABSTRACT

Due to the technological advancement of 3D video technologies and the availability of other supportive services such as high bandwidth communication links, introduction of immersive video services to the mass market is imminent. However, in order to provide better service to demanding customers, the transmission system parameters need to be changed "on the fly". Measured 3D video quality at the receiver side can be used as feedback information to fine tune the system. However, measuring 3D video quality using *Full-reference* quality metrics will not be feasible due to the need of original 3D vide sequence at the receiver side. Therefore, this paper proposed a *Reduced-reference* quality metric for 3D depth map transmission using the extracted edge information. This work is motivated by the fact that the edges and contours of the depth map can represent different depth levels and hence can be used in quality evaluations. Performance of the method is evaluated across a range of Packet Loss Rates (PLRs) and shows acceptable results compared to its counterpart *Full-reference* quality metric.

Index Terms — 3D video transmission, 3D video quality evaluation, *Reduced-reference* quality metric

1. INTRODUCTION

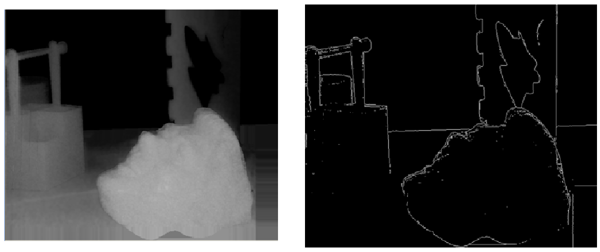
3D video services at home, offices and public places will enhance the way we deal with the technology and enable superior quality immersive video experience. Consumers may expect superior 3D quality at a cost. Therefore, 3D video transmission systems need to be adapted on the fly to the prevailing channel conditions to deliver better quality 3D video. Measured quality at the receiver side can be utilized as feedback information to adjust transmission system parameters. However measuring 3D video quality at the receiver side is a great challenge due to;

- the complex nature of 3D video quality and
- operational difficulties of using *Full-reference* quality metrics where you need to have the original image sequence to measure the quality.

3-D video quality can be described as a collection of different perceptual attributes such as depth perception, presence, eye strain, etc. Therefore, measuring the effect of different system parameter change (e.g. compression level, PLRs) on 3D quality is a complex procedure. Even though several studies are presented on subjective 3D video quality evaluations [1], they have their own limitations and can only be utilized for certain types of image artifacts (e.g. compression artifacts). Due to this diversity and the unavailability of accurate objective quality metric for 3-D video, rigorous and time consuming subjective test campaigns procedures (e.g. ITU-R BT.1438 [2]) are the only feasible method of measuring 3-D video quality at present.

However, researchers have found out that there are strong correlations between subjective ratings and objective quality measures (e.g. VQM, PSNR, SSIM) of individual image components (e.g. colour image and corresponding depth map) of 3D video [3] [4] [5]. This means we can use individual objective quality ratings of 3D video components in place of time consuming subjective test procedures for most of the system parameter changes. However, these candidate objective quality metrics are *Full-reference (FR)* methods [1] [2]. Therefore, the original 3D video sequence should be available at the receiver side to evaluate the quality "on the fly". This is not a viable solution, especially for bandwidth demanding 3D video applications. The only alternative is to find *No-reference (NR)* and/or *Reduced-reference (RR)* quality metrics for 3D video. *No-reference* quality metrics do not need the original sequence at the receiver side to evaluate the quality whereas *Partial-reference* metrics only exploit a subset of the information of the transmitted image in order to perform a comparison. Even though *No-reference* and *Reduced-reference* quality metrics are studied for conventional 2D video (e.g. [6],[7],[8],[9]), studies are not reported for 3D video yet. This paper therefore proposes a *Reduced-reference* quality metric for 3D depth map transmission using their corresponding objective quality measures.

The depth map of a colour plus depth 3D video sequence determines the position of the corresponding colour image in the 3D space (i.e. Depth Image-Based-Rendering/DIBR). Therefore, the quality of the depth maps are crucial since they are used at the receiver side to render novel views [10], [11]. Even though individual pixel values are different, associated depth levels can be represented by the edges or the contour information (i.e. binary edge mask) of the depth maps (see Figure 1). In the proposed *Reduced-reference* quality metric therefore compares the extracted edge information between the distorted image and the original image. More details about the proposed metric with edge detection using filtering are described in Section 2.



(a (b

Figure 1: The *Orbi* (a Depth map and (b Binary edge mask detection using *sobel* filtering

This paper is organized as follows: Section 2 reports the proposed *Partial-reference* quality metric for 3D depth map transmission. This section also investigates the overhead reduction of the proposed method compared to *Full-reference* quality metrics. The experimental setup, results and discussion are presented in Section 3. Section 4 concludes the paper.

2. PROPOSED METRIC

As described in Section 1, edges and contours of the depth maps can represent different depth planes, which decide colour image coordinates in the 3D space. This edge information can be utilized as side information to evaluate quality with a reduced overhead compared to that of *Full-reference* quality metrics. Therefore, this paper proposes a *Reduced-reference* quality metric for 3D depth maps based on edged detection (i.e. binary edge mask as side information) of the depth maps. Edge information of the depth maps can be obtained using different edge detection methods as discussed in Subsection 2.1.

2.1 Edge detection

Edge detection methods can be grouped into two main categories namely gradient and Laplacian [12]. The gradient method detects the edges by looking for the maximum and minimum in the first derivative of the image. The widely used *Sobel* filter also comes under this category of methods. A pixel location is declared an edge location if the value of the gradient exceeds a threshold. Edges will have higher pixel sensitivity values than those surrounding it. Once a threshold is set, the gradient value can be compared to the threshold value and an edge is detected when the threshold is exceeded. In the Laplacian method, edges are detected at the location of zeros in the second derivative. In this work, the *Sobel* operator [13] is selected to obtain edge information (i.e. binary edge mask) due to its simplicity and efficiency.

2.2 Proposed metric

The process for the evaluation of the proposed *Reduced-reference* quality metric using edge information as side information is shown in Figure 2. Initially, side information (i.e. edge information) is generated from the original depth map using *sobel* filtering. This information is then transmitted over the *Reduced-Reference* (RR) channel to the receiver. Ideally, this RR channel should be lossless. In the case of in-band transmission of side information, a high protection through unequal error protection can be provided. At the receiver side the edge information is also obtained from the processed/received depth map. Then these two binary edge masks are compared to obtain a quality index for the proposed method. In this study the case of transmission of depth maps over an IP core network is assumed to evaluate the quality under different Packet Loss Rates (PLRs).

In order to compare the edge information, the commonly used PSNR metric is deployed. However, this comparison is not accurate as comparing the original and processed depth maps (i.e. *Full-reference*) due to the abstract level of information (i.e. binary edge mask) used in the proposed method. Therefore, a relationship is derived between the *Full-reference* and the proposed method using experimental findings. The relationship between these two methods is given by equation (1).

$$PSNR_{Depth\_Map} = f(PSNR_{Binary\_Edge\_Mask}) \qquad (1)$$

$PSNR_{Depth\_Map}$ refers to *Full-reference* PSNR rating for the depth map and $PSNR_{Binary\_Edge\_Mask}$ refers to the PSNR quality rating for the side information (i.e. edge information/Binary edge mask). The derivation of this equation is based on actual experimental results and the details are presented in Section 3.

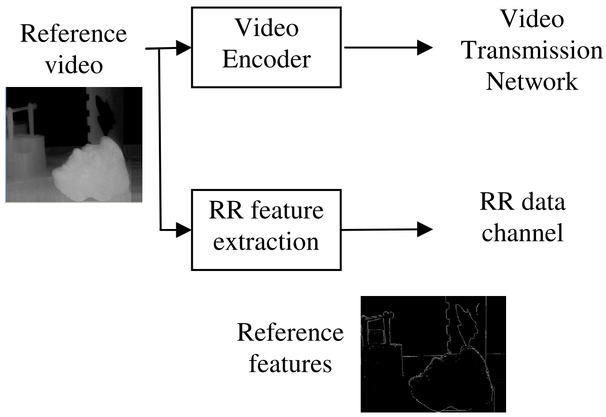


Figure 2: Block diagram of the proposed *Reduced-reference* quality metric (Transmitter side)

However, based on the amount of edge information in the scene the operating range of the $PSNR_{Binary\_Edge\_mask}$ may be different for every sequence. For example, edge map of the *Orbi* sequence is more complex than the *Interview* sequence due to the presence of several foreground objects than in the *Interview* sequence. Therefore, it may be difficult to derive a generic quality metric for all the sequences. In this case, the sequences can be classified based on the expected edge density and lookup tables can be used to evaluate the quality for different sequences. This is the case with different QP values as well. With higher QP values images are smoothed (i.e. removing high frequencies) and detection of edges will be difficult. Hence appropriate edge detection thresholds may necessary. Look up tables can also be used in this case. This allows users to choose the best operating conditions (i.e. different approximations (using Equation (1)) for each sequence category and QP) for the proposed reduced-reference quality metric. In order to evaluate correlations at different quantization levels, experiments are performed for a range of compression levels. Sample lookup table (i.e. R-square values obtained with Equation (1)) for different QPs and sequences is shown in Table 1.

2.3 Overhead analysis

Since reference depth information needs to be transmitted over the channel, either in-band or on a dedicated connection, the overhead should be kept at a minimum level. In case of the proposed method, only the binary image (i.e. ones and zeros of the edge map/1 bit per pixel) will be transmitted and hence require a lower bitrate than the Full-Reference methods (8 bits per pixel).

If we assume 8 bits pixel values for a 720x576 depth map, the Full-Reference method generates 8 x 720 x 576 = 3317760 bits (3.31776 Mbits) per image.

With the proposed method, binary edge mask of the depth map requires 720 x 576 = 414720 bits (414.720 Kbits) per image.

This shows that the proposed method requires less no of bits per reference image than that of the *Full-reference* method. However, still the number of bits required for the reference image is high compared to the compressed bitrates of the depth maps with the latest video coding standards [14]. Therefore, in order to further reduce the overhead, reference data can be compressed, e.g. through run-length encoding. Since the binary edge mask of the depth map is composed of a high number of zero values, a high compression rate is achievable. Moreover, the possibility of selecting a fewer number of blocks from the binary edge mask will be studied in our future work.

3. EXPERIMENTAL SETUP, RESULTS AND DISCUSSION

In order to evaluate the performance of the proposed *Reduced-reference* quality metric for depth maps, experiments are performed for different PLRs (i.e. 3%, 5%, 10% and 20%) and compression levels (i.e. with different QP values). The *Orbi* and *Interview* 3D test sequences (i.e. depth maps of these sequences, 25fps) are encoded using H.264/AVC video coding standard (i.e. JM reference software Version 16.0). Ten seconds long sequences (i.e. 250 frames) are encoded (i.e. with IPPPIPPP… format) using QP values 1, 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50. An I frame is encoded after every 50 frames. Slices (i.e. one MB line = one slice) are also introduced to each frame in order to make the decoding process more robust with the errors. The encoded bit-stream are later simulated over an IP core network using IP error patterns generated for Internet experiments [15]. In order to obtain average results, random starting positions are used for the error pattern files. The corrupted bit-streams with different PLRs are later decoded using JM reference software decoder. *Slice copy* is enabled during decoding to conceal the missing slices of the corrupted bit-stream. At each PLR and QP value, the quality is measured using the PSNR of the depth maps (i.e. *Full-reference* method) and the PSNR of the edge information (i.e. proposed *Reduced-reference* method) generated for the reference image as well as for the processed/corrupted image.

Figure 3 shows the quality measured for the depth maps and the generated binary edge masks with PSNR quality metric at QP=1. It can be clearly see that both measures follow the same trend (i.e. quality is gradually decreasing with the increase of PLR) with an offset. This offset between the $PNSR_{Depth}$ and $PSNR_{Edge}$ quality metrics is due to the reduced amount of side information used in the $PSNR_{Edge}$ method. However, a relationship can be derived between the *Full-reference* (i.e. $PSNR_{Depth}$) and the proposed method ($PSNR_{Edge}$) using a suitable approximation based on experimental data.

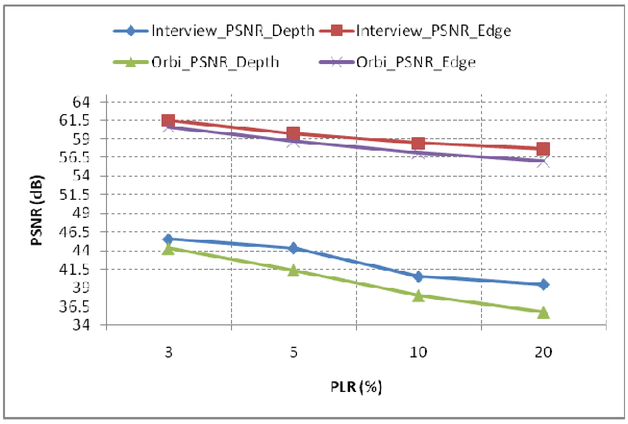


Figure 3: Measured quality using $PSNR_{depth}$ and $PSNR_{edge}$ at QP=1

In order to analyze the correlation between the reference and the proposed methods, the quality ratings are approximated by a second order polynomial (see Equation (2)) for both the sequences at all QP values. Figures 4 and 5 show the scatter plots for the measured image quality using the *Full-reference* and the proposed *Reduced-reference* methods for *Orbi* (QP=35) and *Interview* (QP=45) sequences respectively. Each point of this plot corresponds to the measured average quality using both the *Full-reference* and proposed methods for all the PLRs. According to Figures 4 and 5, it is evident that there is a close relationship between the quality ratings of these methods at all PLRs for a given QP level. Readers should note that these high correlations are achieved for average quality ratings (i.e. averaged over all the frames) for a given PLR and QP.

$$PSNR_{Depth\_Map} = \alpha.PSNR_{Binary\_Depth\_Mask}^{\;2} + \beta.PSNR_{Binary\_Depth\_Mask} + \gamma \qquad (2)$$

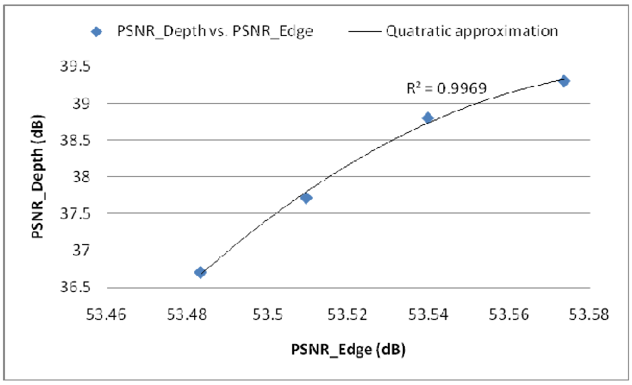


Figure 4: Scatter plot showing the correlation between the proposed *Reduced-reference* method and the reference method (*Orbi*, QP=35).

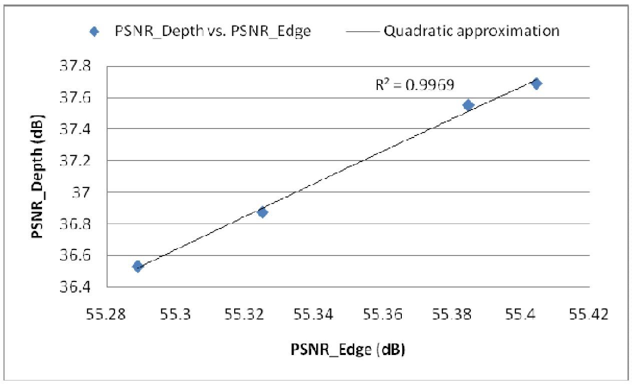


Figure 5: Scatter plot showing the correlation between the proposed *Reduced-reference* and the reference methods (*Interview*, QP=45).

Individual correlation coefficients of determination (i.e. R-square) for *Orbi* and *Interview* sequences at different QP levels are listed in Table 1. According to Table 1, it is clear that proposed method is highly correlated with the *Full-reference* method for all the QP values. This suggests that we can use the proposed method in place of the *Full-reference* method with a reasonable accuracy at all PLRs.

Table 1: Quality ratings for different QP values

QP	R-square		QP	R-square	
	Orbi	**Interview**		**Orbi**	**Interview**
1	0.9993	0.9678	30	0.9870	0.9977
5	0.9822	0.9860	35	0.9969	0.9901
10	0.9928	0.9937	40	0.9978	0.9539
15	0.9995	0.9998	45	0.9969	0.9969
20	0.9999	0.9158	50	0.8415	0.9524
25	0.9711	0.9770	-	-	-

However it is difficult to obtain a more generic approximation for both *Orbi* and *Interview* sequences. This is due to the different level of edge information in these sequences. Therefore generated lookup tables (e.g. Table 1 shows the R-square values obtained with quadratic approximation/Equation (2)) with approximated correlations can be utilized to measure the quality of different categories of scenes (i.e. scenes can be categorized based on the edge density). Lookup table provides the exact relationship (i.e. α, β and γ values of Equation (2)) for a given category of sequence and QP level.

Figures 6 and 7 illustrate the effectiveness of the proposed method for measuring depth map quality in the absence of the original depth map images for *Orbi* and *Interview* sequences respectively. According to these figures the quality ratings of the proposed method (approximated with Equation (2)) are close to the *Full-Reference* method at all PLRs.

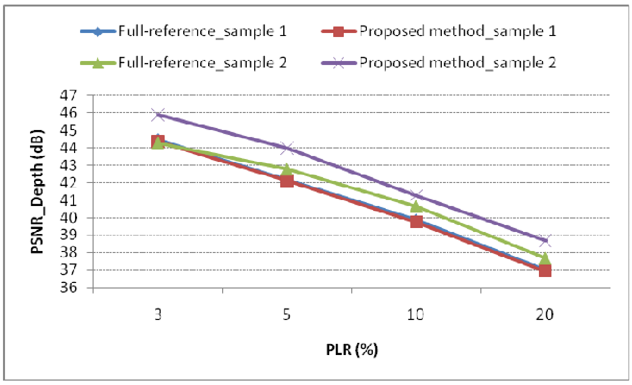


Figure 6: Measured quality for the *Orbi* at different PLRs (QP=20, Equation (2): α=4.864, β=-503.77 and γ=130.65)

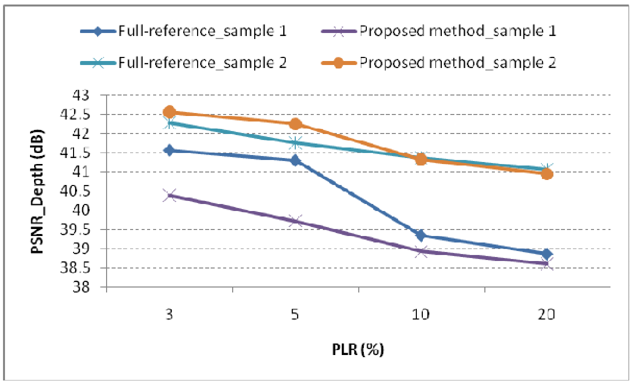


Figure 7: Measured quality for the *Interview* at different PLRs (QP=35, Equation (2): α=30.727, β=-3376.3 and γ=92786).

4. CONCLUSION

This paper proposed a *Reduced-reference* quality metric for 3D depth map transmission based on edge detection. Since edges of the depth map represent different depth levels, edge information is used as the side information in the proposed method. Since edge information only contains binary zeros and ones, this method requires less bandwidth compared to the *Full-reference* method. PSNR metric is utilized to compare the edge information (i.e. obtained binary edge masks). In order to obtain matched results, the proposed method is calibrated against the *Full-reference* PSNR metric. Results show good approximation for the *Full-reference* quality metric for all considered PLRs. Lookup tables generated for different scene categories and QP levels can be effectively used in measuring quality with the proposed method. This suggests that due to the practical problems associated with *Full-reference* method (i.e. bandwidth), *Reduced-reference* quality metrics as described in this paper is an acceptable compromise for the 3D video research and development community.

5. ACKNOWLEDGEMENT

This work was partially supported by the European Commission (FP7 project OPTIMIX INFSO-ICT-21462).

6. REFERENCES

[1] P. Seuntians, L. Meesters, and W. IJsselsteijn, “Perceptual evaluation of JPEG coded stereoscopic images”, *Proc. SPIE*, vol.5006, pp. 215-226, 2003.

[2] Int. Telecommunication Union/ITU Radio Communication Sector, “Subjective assessment of stereoscopic television pictures”, ITU-R BT.1438, Jan. 2000.

[3] 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", IEEE Journal of Selected Topics in Signal Processing-Special Issue on Visual Media Quality Assessment, Vol. 3, No. 2, pp. 304-318, 2009.

[4] C.T.E.R. Hewage, S.T. Worrall, S. Dogan, and A.M. Kondoz, "Prediction of Stereoscopic Video Quality Using Objective Quality Models of 2-D Video", IET Electronics Letters, Vol. 44, No. 16, pp. 963-965, 31st July 2008.

[5] S.L.P. Yasakethu, C.T.E.R. Hewage, W.A.C. Fernando, S.T. Worrall and A.M. Kondoz, "Quality Analysis for 3D Video Using 2D Video Quality Models", IEEE Trans. on Consumer Elec., Vol. 54, No. 4, pp. 1969-1976, Nov. 2008.

[6] R. V. Babu, A. S. Bopardikar, A. Perkis, O. I. Hillestad, “No-Reference Metrics For Video Streaming Applications”, Proc. International Packet Video Workshop 2004.

[7] Z. Wang, H. R. Sheikh and A. C. Bovik, “No-Reference Perceptual Quality Assessment of JPEG Compressed Images”, Proc. IEEE Int. Conference on Image Processing, pp. 477-480, Rochester, New York, September 22-25, 2002.

[8] Z. Wang, E. Simoncelli, “Reduced-reference image quality assessment using a wavelet-domain natural image statistic model”, 17th SPIE annual symposium on electronic imaging, San Jose, CA, Jan 2005.

[9] S. Wolf, M. H. Pinson, “Low Bandwidth Reduced Reference Video Quality Monitoring System”, First International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, Arizona, Jan. 2005.

[10] C.T.E.R. Hewage, S. Worrall, S. Dogan, H. Kodikara Arachchi, and A.M. Kondoz, "Stereoscopic TV over IP", *Proceedings of the 4th IET European Conference on Visual Media Production (CVMP 2007)*, pp. 1-7, London, UK, Nov. 2007.

[11] A. Tikanmäki, A. Gotchev, A. Smolic, K. Müller, "Quality Assessment of 3D Video in Rate Allocation Experiments", *Proc. IEEE International Symposium on Consumer Electronics (ISCE'08)*, Algarve, Portugal, April 2008.

[12] D. Marr and E. Hildreth, “Theory of edge detection,” in Proce. of the Royal Society of London. Series B, 1980.

[13] J. Woods, Multidimensional Signal, Image and Video Processing and Coding, Elsevier, 2006.

[14] C.T.E.R. Hewage, H.A. Karim, S. Worrall, S. Dogan, and A.M. Kondoz, "Comparison of Stereo Video Coding Support in MPEG-4 MAC, H.264/AVC and H.264/SVC", Proceedings of the 4th IET Intern. Conf. on Visual Information Engineering (VIE 2007), London, UK, 25-27 July 2007.

[15] S. Wenger. “Error patterns for Internet video experiments”, *ITU-T SG16 Document Q15-I-16-R1*, 1999.