REDUCED-REFERENCE QUALITY EVALUATION FOR COMPRESSED DEPTH MAPS ASSOCIATED WITH COLOUR PLUS DEPTH 3D VIDEO

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

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

ABSTRACT

3D video quality assessment is a key task in transmission of immersive video, both for the assessment of end-to-end system performance and for quality evaluation "on the fly". For example, measured 3D video quality at the receiver side can be used as feedback information for controlling transmission parameters. However, measuring video quality at the receiver side using *Full-Reference* quality metrics (e.g. PSNR) is less practical due to the need of original image sequence at the receiver. Hence, *Reduced-Reference* and *No-Reference* quality metrics are being developed to evaluate the quality. Since edges and contours of the depth map represent different depth levels, this paper proposes a *Reduced-Reference* quality metric for the depth maps associated with colour plus depth 3D video using edge detection. Edge/Side information (i.e. binary edge mask) generated using *sobel* filtering is utilized. Results show that the proposed metric can be used in place of a *Full-Reference* metric with an acceptable accuracy.

Index Terms— Colour plus depth video, *Reduced-Reference* quality metric, Edge detection, *Sobel* filtering

1. INTRODUCTION

3-D video quality can be described as a collection of different perceptual attributes such as depth perception, presence, eye strain, etc. Due to this diversity and the unavailability of accurate objective quality metrics for 3-D video, rigorous and time consuming subjective test campaigns 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 [1] [2] [3]. This means we can use individual objective quality ratings of 3D video components in place of time consuming subjective test procedures for every system parameter change.

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 *Reduced-Reference* metrics only exploit a subset of the information of the transmitted image in order to perform a comparison.

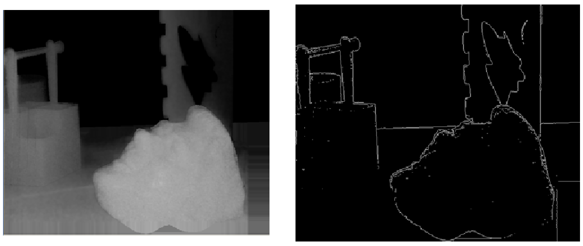
A number of studies focusing on *No-Reference* and *Reduced-Reference* quality metrics for 2D video can be found in the Literature. For instance, [4] [5] present *No-Reference* quality metrics for conventional video. However, these metrics are designed for a specific type of distortions such as *blurriness* and *blockiness* and often fail in providing a reliable assessment. *Reduced-Reference* quality metrics are the next favorite due to more accurate measures compared to *No-Reference* methods with the cost of an overhead. For example, *Reduced-Reference* quality metrics based on a wavelet domain model and NTIA General Video Quality Model (VQM) are presented in [6] and [7] respectively. *No-Reference* or *Reduced-Reference* quality metrics for 3D video quality evaluations are not reported yet. However, *No-Reference* and *Reduced-Reference* quality metrics proposed for 2D video can be individually applied for each view of a multi-view 3D video sequence. In the case of colour plus depth map 3D video, aforementioned 2D metrics can also be applied to colour and depth map sequences separately, even though the colour and depth map characteristics and their usage are different.

Edges and contours of the depth maps represent different depth planes, which decide colour image coordinates in the 3D space. Therefore, this paper proposes a *Reduced-Reference* quality metric for depth maps based on edge detection (i.e. binary edge mask as side information) of the depth maps. More details about the proposed metric with edge detection using different filtering methods are described in Section 2.

This paper is organized as follows: Section 2 reports the proposed *Reduced-Reference* quality metric for depth maps. 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 REDUCED-REFERENCE METRIC

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 [1] [3]. Even though individual pixel values are different, associated depth levels can be represented by the edges or the contour information of the depth maps (see Figure 1). The proposed *Reduced-Reference* quality metric therefore compares the edge information between the distorted image and the original image.



(a) (b)

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

2.1. Edge detection

Edge detection methods can be grouped into two main categories namely gradient and Laplacian [8]. 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 [9] is selected to obtain edge information (i.e. binary edge mask) due to its simplicity and efficiency.

2.2. Proposed method

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 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 lossless transmission of depth maps is assumed in order to evaluate the quality under image compression artifacts.

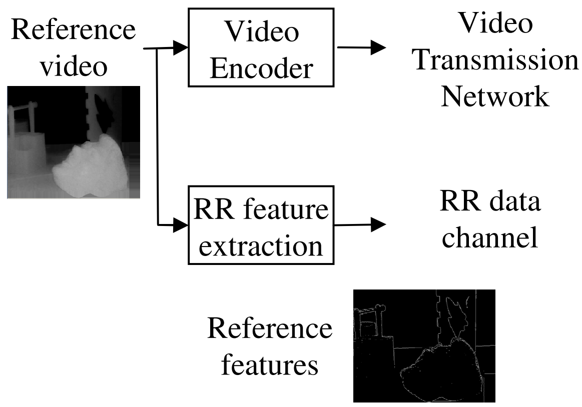


Figure 2: Block diagram of the proposed *Reduced-Reference* quality metric (Transmitter side)

In order to compare the edge information, the commonly used PSNR metric is deployed. However, this comparison is not accurate as comparing the original depth map and the processed one (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).

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

$PSNR_{Depth\_Map}$ referes 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.

2.3. Overhead

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-bit 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 number 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 [10]. 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 a range of 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. single layer configuration of JSVM reference software Version 9.12). Ten-second long sequences (i.e. 250 frames) are encoded (i.e. with IPPP… format) using QP values 1, 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50. At each 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. *Reduced-Reference* method) generated for the reference image as well as for the processed image.

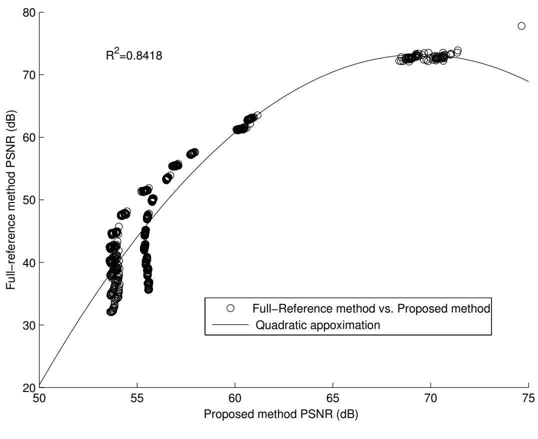


Figure 3: Scatter plot showing the correlation between the proposed *Reduced-Reference* method and the reference method.

In order to obtain quality ratings as accurate as the *Full-Reference* quality metric, the proposed metric has to be set and calibrated. Therefore, the correlation between the reference and the proposed methods is analyzed. Figure 3 shows the scatter plot for the measured image quality using the *Full-Reference* and the *Reduced-Reference* methods for both *Orbi* and *Interview* sequences. Each point of this plot corresponds to the measured quality using both the *Full-Reference* and proposed methods for a specific QP value. According to Figure 3, it is evident that there is a close relationship between the quality ratings of these methods at all QP levels. Individual correlation coefficient of determination for *Orbi* and *Interview* sequences are 0.89 and 0.86 respectively. Overall coefficient for both sequences is as high as 0.84. However, it can be observed that the proposed method is less sensitive to image artifacts at higher QP values. This may be partly due to the removal of high frequency information from the depth map images at high QP values. As a result the edge detection threshold may have failed to detect many edges on these images.

The relationship between these two quality metrics (i.e. *Full-Reference* PSNR and the *Reduced-Reference* PSNR) is approximated by a second order polynomial. In the case of 720 x 576 resolution depth maps, the following normalization has been chosen (see Equation (2)).

$$PSNR_{Depth\_Map} = -\alpha \times PSNR^2{}_{Binary\_Depth\_Mask} + \beta \times PSNR_{Binary\_Depth\_Mask} - \gamma \qquad (2)$$

with α = 0.1396, β = 19.385 and γ = 599.89. These α, β and γ values are determined based on correlation analysis (see Figure 3). In other cases, higher order polynomials may be necessary to maximize the performance.

Table 1 shows the quality measured with both the reference and proposed methods. It can be clearly seen that the proposed method most of the time matches the quality measured by the *Full-Reference* method. Figures 4 and 5 also 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 are close to the *Full-Reference* method at low QP values. However, the performance of the *Reduced-Reference* method is lower at high QP values. This is mainly due to the weaker performance of the edge detection algorithm at high QP values. Since high frequency image information which can be used to detect edges is eliminated during the encoding process, the proposed method does not perform well at high compression ratios (i.e. with larger QP values).

Table 1: Quality ratings for different QP values

QP	*Interview*		*Orbi*	
	Reference	**Proposed**	**Reference**	**Proposed**
1	72.764523	73.030653	72.936808	72.80031
5	62.935555	62.612877	61.346371	61.47653
10	57.374113	54.195689	55.451189	51.26707
15	53.393386	49.817805	51.441996	45.61736
20	49.934554	47.045574	47.662817	41.01952
25	47.243582	46.087898	44.674246	39.02034
30	44.718597	45.630225	42.451183	38.68711
35	42.449527	45.512658	40.456503	38.76760
40	40.279798	45.874401	38.561374	39.03393
45	38.417205	46.117952	36.240363	39.20856
50	36.256412	46.314145	33.558658	39.24268

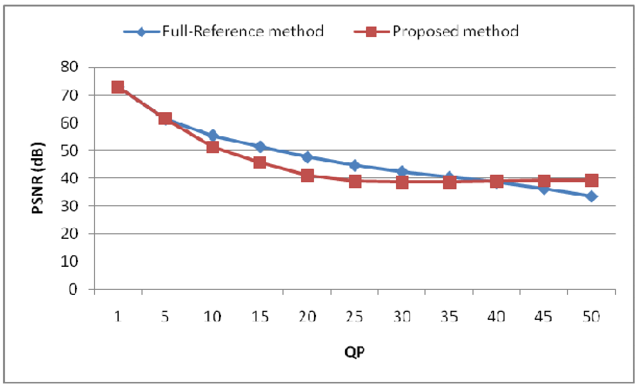


Figure 4: Measured quality for the *Orbi* at different QP values

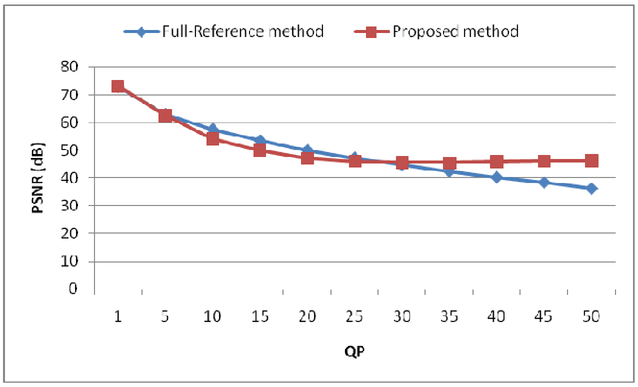


Figure 5: Measured quality for the *Interview* at different QP values

4. CONCLUSION

Quality assessment for emerging 3D video services "on the fly" will be a great challenge due to the complex nature of 3D video quality and less possibilities to use a *Full-Reference* method. Hence, this paper proposed a *Reduced-Reference* quality metric for compressed depth maps associated with colour plus depth 3D video based on edge detection. Since edges of the depth map represent different depth levels, edge information is used as the side information in this method. This method requires less bandwidth compared to the *Full-Reference* method as edge information only contains binary ones and zeros. The comparison of edge information (binary edge mask) is done using PSNR metric. Results show good approximation for the *Full-Reference* quality metric, especially at higher bitrates. 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] 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, April 2009.

[2] 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.

[3] 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 Transactions on Consumer Electronics*, Vol. 54, No. 4, pp. 1969-1976, November 2008.

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

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

[6] Z. Wang, E. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model", *17th SPIE annual sym. electronic imaging*, CA, Jan 2005.

[7] 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, January 23-25, 2005.

[8] D. Marr and E. Hildreth, "Theory of edge detection," in *Proceedings of the Royal Society of London. Series B*, 1980.

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

[10] 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 International Conference on Visual Information Engineering (VIE 2007)*, London, UK, 25-27 July 2007.