

Prediction of stereoscopic video quality using objective quality models of 2-D video

C. T. E. R. Hewage, S. T. Worrall, S. Dogan and A. M. Kondoz

The timely deployment of 3-D video applications requires accurate objective quality measures, so that time consuming subjective tests can be avoided. Investigates is the correlation between subjective and objective evaluations of colour plus depth map 3-D video. The investigation is conducted for different compression ratios. Subjective tests are performed to determine the overall image quality and depth perception of a range of asymmetrically coded video sequences. The subjective results are used to determine more accurate objective quality assessment metrics for colour plus depth map based stereoscopic video.

Introduction: In the near future, many conventional 2-Dimensional (2-D) video applications are likely to be replaced by immersive video to provide a sense of “being there”. Even though the technologies required for 3-D video are emerging rapidly, the effect of these technologies on the perceptual aspects of 3-D viewing has not been thoroughly studied. Furthermore, the accuracy of objective measures for 3-D video has not been investigated, which means that there is currently a need for time consuming and rigorous subjective evaluation tests. In return, this significantly hinders the wide-spread proliferation of 3-D video applications into the consumer market.

Objective quality models, such as Video Quality Model (VQM) [1], are emerging to allow more accurate measurement of the quality of 2-D video

compared to the commonly used objective measures (e.g. PSNR), which are not closely correlated with the characteristics of the Human Visual System (HVS). However, the development of an objective quality metric, which incorporates the perceptual aspects of 3-D, is a complex process as the perceived attributes (e.g., depth, presence, naturalness, visual comfort) of 3-D are multi-dimensional in nature. Seuntings et al describes in [2] that JPEG coding of stereoscopic video has an effect on overall image quality, but no effect on perceived depth. In [3], the added value of depth is not taken into account when assessing the perceived image quality of MPEG-2 coded stereo sequences. In contrast, a positive relationship between depth and perceived image quality for uncompressed stereoscopic images is shown in [4].

In this letter, subjective evaluation testing is used to measure the effect of compression artifacts on the perceived quality (i.e. overall image quality and depth perception) of the reconstructed stereoscopic video. Furthermore, the relationship between the perceived overall image quality and the perceived depth are also discussed. Finally, the perceived attributes of reconstructed stereoscopic video are predicted using three objective quality models namely PSNR, VQM [1] and SSIM (Structural SIMilarity) [5]. The results help us answer important questions concerning the correlation between objective and subjective measurements of 3-D video quality, which in return allows us to identify the most adequate methods for objective measurement of the quality of 3-D video.

This investigation focuses on the colour plus depth representation of stereoscopic video, which is widely used in research and standardization

activities [6], [7], due to its flexible representation and adaptability for existing coding and transmission technologies. The per-pixel depth value determines the position of the associated colour texture in the 3-D space. The image warping technique known as Depth-Image-Based Rendering (DIBR) is used to render the desired binocular viewpoint image [7].

Method: In order to meet the storage and bandwidth requirements of stereoscopic video applications, efficient compression techniques are required. Existing 2-D compression methods can be adopted for encoding of 3-D video [7]. In this letter, a layered coding configuration, based on the scalable extension of H.264/AVC is chosen to encode colour and depth map video, due to its adaptability and good compression efficiency [9].

The effect of different artifacts introduced by image coding on the perceived quality of 3-D video may be diverse in nature [2]. Therefore, appreciation oriented psychophysical scaling paradigms can be used in the assessment of emerging 3-D technologies, such as stereoscopic TV and stereoscopic video streaming. In this explorative study, subjective tests are performed in order to obtain the perceived stereoscopic video quality for different combinations of symmetric/asymmetric coding. The subjective image quality ratings are mapped against the candidate objective measures using the symmetric logistic function, described in ITU-Recommendation BT.500-11 [10], which specifies a methodology for carrying out subjective quality assessments for TV pictures. The objective measures of colour image sequence and rendered left and right views using the DIBR technique, are used to predict the perceived 3-D video quality.

Experiment and Results: Four colour and depth based stereoscopic video sequences, namely *Orbi*, *Interview*, *Break dance* and *Ballet*, are used in the experiments. The *Orbi* and *Interview* sequences (720x576 pixels and 25 frames/second) are captured using a depth/range camera whereas the *Break dance* and *Ballet* depth maps (1024x768 pixels and 15 frames/second) are obtained from the multi-view image sequences, via stereo matching techniques, by the Interactive Visual Media group at Microsoft Research [8].

The JSVM (Joint Scalable Video Model) reference software version 9.4 is used to encode each test sequence at different average bitrates. Video sequences of 10 seconds are encoded using the IPPP... sequence format, and Context Adaptive Binary Arithmetic Coding (CABAC). The Quantisation Parameters (QPs) of the base and enhancement layers are varied to obtain symmetrically/asymmetrically coded colour and depth image sequences. In order to highlight the visibility of coding artifacts, the QP sizes 30, 35, 40, 45, and 50 are selected for both colour and depth image sequences. The coding combinations 30_30, 30_35, 30_40, 30_45, 30_50, 35_30, 35_35, 40_30, 40_40, 45_30, 45_45, 50_30, 50_50 are used in the experiments as a means for reducing the bitrate requirements for colour plus depth map based stereoscopic video. The first and second numbers of these combinations refer to the colour and depth map QPs used during encoding, respectively.

A 42" Philips WOWvx multi-view auto-stereoscopic display is used in the experiments to display the coded stereoscopic image sequences. The 3-D display is calibrated using a GretagMacbeth Eye-One Display 2 calibration device. During the subjective evaluation tests, observers are asked to rate the

video sequences according to the Double Stimulus Continuous Quality Scale (DSCQS) method, as described in ITU-Recommendation BT.500-11 [10]. The perceived overall image quality and perceived depth are rated on a quality categorical scale from 1 to 5, where 1 represents bad image quality/depth perception and 5 represents excellent image quality/depth perception. Furthermore, the scale is labeled with the adjective terms namely [bad], [poor], [fair], [good], and [excellent] according to [10].

During the analysis of the results, the difference in subjective ratings for the impaired image sequences and the original image sequences are calculated. Then the difference is scaled into a linear opinion score scale, which ranges from 0 (excellent: same quality as the original sequence) to 100 (bad: much worse quality than the original sequence). Then the Mean Opinion Score (MOS) for each test sequence is obtained after averaging the opinion scores for all subjects.

Thirty two non-expert observers (ten female and twenty two male observers) participated in the experiments. They are divided into two groups in order to assess perceived overall image quality and perceived depth. All participants have a visual acuity of ≥ 1 (as tested with the Snellen chart), good stereo vision < 60 seconds of arc (as tested with the TNO stereo test), and good colour vision (as tested with the Ishihara test). The stimulus set contains thirteen coded video sequences and the original, uncompressed version of each scene is used as the reference in the evaluation test. All combinations of the coded colour and depth video sequences are presented twice. Therefore, in total four test sequences, one repetition and 13 coding combinations are used. This resulted in a stimulus set of $4 \times 2 \times 13 = 104$ video sequences. A

set of 104 stereoscopic video sequences is randomised and presented sequentially.

Fig. 1 shows the MOS scores for the overall perceived image quality, depth perception, the objective colour image quality measured using VQM, and the objective quality of the average of the rendered left and right views measured using VQM. The bitrates increase from the left side of the graph to the right. Furthermore, the Standard Error (SE) for the perceived image quality/depth and the 5% error margin for VQM scores are also presented.

The subjective quality scores for image quality increase as bitrates increase, for all of the tested sequences. But there are visible quality dips when the coding configuration contains high QP values for the colour video component (e.g., 50_50, 45_45, 50_30, and 45_30). These quality dips prevent the formation of a straightforward relationship between perceived overall image quality and average bitrates. However, medium QP values (i.e., 35 and 40) for colour video coding, together with any QP value for depth video coding can be used to achieve good perception of overall image quality and depth. This means that depth can be coded with a higher QP value to reduce the bitrate, without compromising the overall perceived image quality. The VQM score of the colour video and average VQM score of the rendered left and right video, using coded colour and depth sequences, demonstrate similar trends compared to the perceptual image quality scores, but with an offset in terms of the MOS score. This suggests that the addition of the depth increases viewer satisfaction significantly, masking the effects of compression artifacts in the colour video. In addition to that, the VQM ratings of the colour and

rendered left and right video could be used as a good objective quality model for prediction of the perceived image quality and depth.

The depth perception shows a similar trend in image quality perception, but on average has lower MOS scores compared to the overall image quality results. Therefore, depth perception is not significantly affected by using different QP values for depth map coding, whereas the artifacts (i.e., blocking, blurring) introduced to the colour video by compression heavily influence depth perception (e.g., the depth perception is low for the 35_30, 40_30, 45_30, and 50_30 coding combinations). Furthermore, it can be deduced from Fig. 1 that there is high correlation between the overall image quality and the depth perception. The correlation coefficient of determination between perceived image quality and depth ratings is 0.92. Thus, the perception of depth is positively influenced by the perceived image quality for coded colour and depth sequences. Therefore, objective measures of depth video are not used in predicting the perceived 3-D video quality.

The objective quality assessment models used in predicting perceived quality attributes of 3-D are PSNR/SSIM/VQM scores of the coded colour video, and the average PSNR/SSIM/VQM of the rendered left and right views, generated using the coded colour plus depth sequences. The quantitative measures for each prediction model, approximated using the symmetrical logistic function described in [10], are presented in Table 1. This shows that the VQM scores of the colour video sequence is a good model for predicting both the perceived image quality and the depth. Furthermore, the average VQM of the rendered left and right view shows high correlation with the perceived depth of coded stereoscopic video.

Conclusion: This letter has addressed the lack of adequate 3-D video objective quality assessment metrics, which could hinder the further development and introduction of new 3-D video services. The results show that the output from the VQM objective metric can be mapped, so that it correlates strongly with both the overall viewer perception of image quality and depth perception. This suggests that, while subjective test results should remain the “gold standard”, the use of objective quality assessment metrics, as described in this letter, is an acceptable compromise for the 3-D video research community.

References

- 1 Pinson, M., and Wolf, S.: ‘A new standardized method for objectively measuring video quality’, IEEE Transactions on Broadcasting, September 2004, Vol. 50, pp. 312-322
- 2 Seuntiëns, P., Meesters, L., and IJsselsteijn, W.: ‘Perceptual evaluation of JPEG coded stereoscopic images’, Proc. SPIE, 2003, Vol. 5006, pp. 215-226
- 3 Tam, W., Stelmach, L., and Corriveau, P.: ‘Psychovisual aspects of viewing stereoscopic video sequences’, Proc. SPIE, 1998, Vol. 3295, pp. 226-235
- 4 IJsselsteijn, W., Ridder, H. de., and Vliegen, J.: ‘Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration’, IEEE Trans. Circ. Syst. Video Tech., 2000, Vol. 10, pp. 225-233
- 5 Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P.: ‘Image Quality Assessment: From Error Visibility to Structural Similarity’, IEEE Transactions on Image Processing, 2004, Vol. 13, pp. 600-612
- 6 Text of ISO/IEC FDIS 23002-3 Representation of Auxiliary Video and Supplemental Information: ISO/IEC JTC1/SC29/WG11, Marrakesh, Morocco, January 2007, Doc. N8768
- 7 Fehn, C.: ‘Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV’, Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems XI, 2004, pp. 93-104

8 Zitnick, C.L., Kang, S.B., Uyttendaele, M., Winder, S., and Szeliski, R.: 'High-quality video view interpolation using a layered representation', ACM SIGGRAPH and ACM Trans. on Graphics, Los Angeles, 2004, pp. 600-608

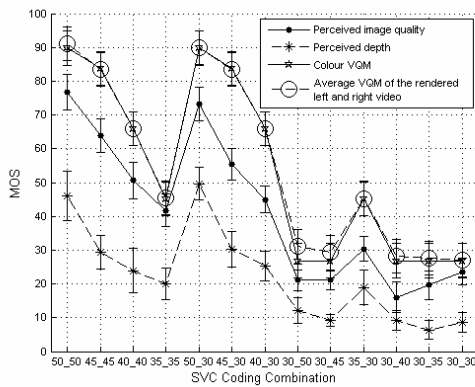
9 Hewage, C.T.E.R., Karim, H.A., Worrall, S., Dogan, S., and Kondoz, A.M.: 'Comparison of Stereo Video Coding Support in MPEG-4 MAC, H.264/AVC and H.264/SVC', In Proc. of 4th IET conf. on VIE, London, UK, 2007.

10 International Telecommunication Union (ITU) Radio communication Sector.: 'Methodology for the subjective assessment of the quality of television pictures', ITU-R BT.500-11, January 2002.

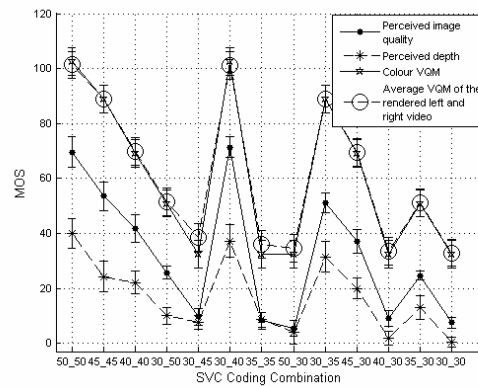
Authors' affiliations:

C.T.E.R. Hewage S. Worrall S. Dogan and A.M. Kondoz (I-Lab, CCSR, Faculty of Engineering and Physics, University of Surrey, Guildford, GU2 7XH, Surrey, United Kingdom; E-Mail: e.thushara@surrey.ac.uk)

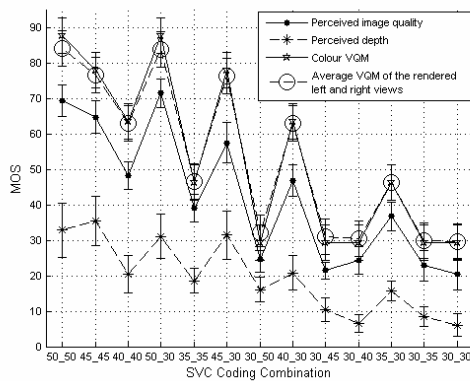
Fig. 1 The MOS scores for the perceived image quality, depth perception, colour measured using VQM, and the average of the rendered left and right views measured using VQM: (a) Orbi (b) Interview (c) Break dance (d) Ballet



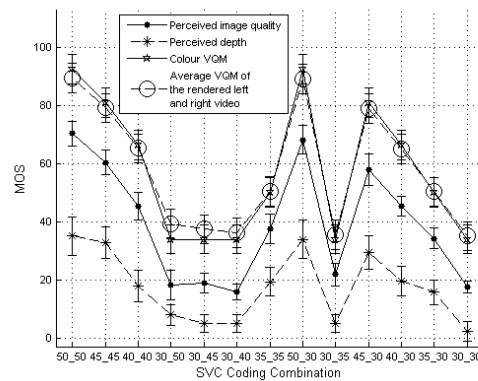
(a)



(b)



(c)



(d)

Table 1 Performance of image quality assessment models. CC: Correlation Coefficient; RMSE: Route Mean Squared Error; SSE: Sum of Squares due to Error

Objective quality model	Overall image quality			Depth perception		
	CC	RMSE	SSE	CC	RMSE	SSE
Colour PSNR	0.8100	0.0924	0.4265	0.8192	0.0529	0.1401
Average PSNR of the Rendered Left and Right views	0.7356	0.1090	0.5936	0.7768	0.0588	0.1729
Colour VQM	0.8904	0.0701	0.2460	0.8421	0.0490	0.1224
Average VQM of the Rendered Left and Right	0.8764	0.0745	0.2774	0.8432	0.0493	0.1215
Colour SSIM	0.7618	0.1034	0.5347	0.7274	0.0650	0.2112
Average SSIM of the Rendered Left and Right Views	0.7528	0.1054	0.5550	0.7195	0.0659	0.2173