A DISOCCLUSION REPLACEMENT APPROACH TO SUBJECTIVE ASSESSMENT FOR DEPTH MAP QUALITY EVALUATION

N. Haddad[1], S. Dogan[1], H. Kodikara Arachchi[1], V. De Silva[2], and A.M. Kondoz[3]

[1]I-Lab/CVSSP, University of Surrey, Guildford, UK; [2]Apical Limited, Loughborough, UK;
[3]Institute for Digital Technologies, Loughborough University in London (LUiL), London, UK
[1]{nasser.haddad, s.dogan, h.kodikaraarachchi}@surrey.ac.uk,
[2]varunax@gmail.com, [3] a.kondoz@lboro.ac.uk

ABSTRACT

An inherent problem of Depth Image Based Rendering (DIBR) is the visual presence of disocclusions in the rendered views. This poses a significant challenge when the subjective assessment of these views is utilised for evaluating the quality of the depth maps used in the rendering process. Although various techniques are available to address this challenge, they result in concealing distortions, which are directly caused by the depth map imperfections. For the purposes of depth map quality evaluation, there is a need for an approach that deals with the presence of disocclusions without having further impact on other distortions. The aim of this approach is to enable the subjective assessments of rendered views to provide results, which are more representative of the quality of the depth map used in the rendering process.

Index Terms **—** 3DTV, depth map, disocclusions, mask replacement, subjective evaluation.

1. INTRODUCTION

The role of depth information within emerging 3D video applications (such as 3DTV) has increased during recent years. For example, the effect of the accuracy of a depth map on the quality of virtually synthesised views is an integral part in providing good Quality of Experience (QoE) levels, within these emerging applications. Therefore, the concept of the quality of depth maps has been subjected to ever increasing scrutiny.

Synthesised virtual (rendered) views are generated through a Depth Image Based Rendering (DIBR) process. A major issue in DIBR is that certain regions of a rendered view may not be visible within the view used to generate them. This issue is commonly referred to as “disocclusions”. Such disoccluded areas are visually prominent when the rendered view is being extrapolated. In other words, the target rendered view’s location lies outside the existing camera baseline [1].

Disocclusions (also known as “holes”) have a significant impact on the subjective assessment of the rendered views. This impact is of extra importance when the subjective assessment results are used to evaluate the quality of the depth maps utilised in the rendering process. The prominent visual presence of the disocclusions within the rendered view outweighs the presence of other artefacts resulting from lower quality depth maps. This causes misrepresentative subjective results; therefore, subjective assessment of rendered views becomes an ineffective tool for evaluating the quality of the depth maps used in the rendering process.

Disocclusions within the rendered view have to be filled adequately; otherwise, annoying artefacts appear in the disoccluded regions within the rendered views [2]. Work carried out towards solving the issue of the disocclusion areas is extensive, with varying methods in addressing this problem. One approach to solve the disocclusion issue is the pre-processing of depth information by applying filtering techniques, so that no disocclusions occur in the rendered views. Although this method eliminates the presence of disocclusions in the rendered view, it also introduces geometrical distortions within those views [3].

Another approach to address the disocclusion issue is covering those regions with suitable, known image information. This approach is known as hole filling. Several hole filling techniques are available, with varying degrees of computational complexity and different degrees of impact on the quality of the rendered views [3], [4].

The approaches suggested in the literature to handle the disocclusion issue also have an impact on the subjective evaluation of the rendered images. Despite the fact that they are aimed at addressing the disocclusion problem, they also result in concealing other artefacts or distortions caused by inaccuracies in the depth information. In turn, this results in misrepresentative subjective results, thus rendering subjective assessment ineffective with respect to evaluating the quality of depth maps used in the rendering process.

This paper presents a proposed approach targeted at concealing the disocclusion regions, without impacting the presence of other distortions and artefacts within the rendered view. The aim of this approach is to enable the subjective assessments of rendered views to provide results, which are more representative of the quality of the depth map used in the rendering process. The impact of this approach on the subjective assessment results is examined, utilising well-established objective measurements. The proposed approach is also applied within the objective measurement procedure for further examination.

The remainder of this paper is organised as follows. The proposed method is described in Section 2. The experimental setup is explained in Section 3. Section 4 presents the experimental results and discussion. Finally, Section 5 concludes the paper along with pointers to future work.

2. MASK REPLACEMENT APPROACH

The idea behind the proposed approach revolves around the replacement of the disocclusion regions with suitable information, without eliminating or concealing any other distortions within the rendered view. The identification of the disocclusion regions is achieved by utilising the occlusion layer “hole mask” feature present within the MPEG software View Synthesis Based on Disparity/Depth (ViSBD) version 2.1 [5]. This software has been utilised for both view rendering and hole filling purposes in this paper.

The generated rendered views are selected to be at the location of an already existing colour view. The hole mask generated by the ViSBD software is subsequently used to extract the corresponding colour information related to the disoccluded regions from the original view. The extracted colour information is then used to fill in the disoccluded regions in the targeted rendered view to generate a masked rendered view [6]. Resulting masked views were utilised for the assessment purposes. The mask replacement process is illustrated in Figure 1.

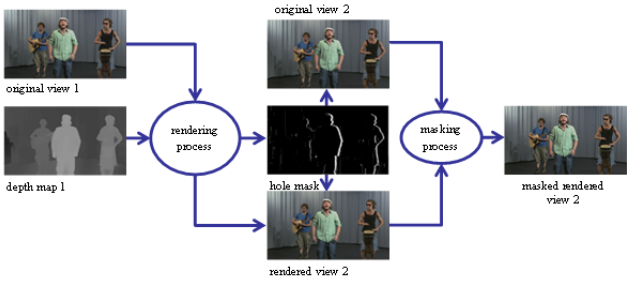


Figure 1. Disocclusions mask replacement process.

3. EXPERIMENTAL SETUP

The evaluation of this proposed disocclusion replacement approach includes the assessment of both the hole filled (non-masked) rendered views and the views generated using the proposed approach (masked views). The rendered views are generated utilising, readily available, non-pre-processed depth maps and compressed versions of these depth maps. The assessment process involves both subjective and objective evaluations of image and video stimuli obtained from both masked and non-masked rendered views. The following subsections detail the assessment procedure and the dataset utilised.

3.1 Dataset

Six original Multi-View plus Depth (MVD) sequences (namely: *Band, BMX, Musicians, Poker [7], Act and Parisband [8]*), were considered for the assessment procedure. The selected sequences cover a range of texture complexities and variable motion content. All the sequences used are of 1920x1080 pixels resolution and were captured at a frame rate of 25 fps and are 10 seconds in duration.

The available depth maps for the utilised sequences were compressed at three different levels. Quantization Parameters (QPs) were selected at 22, 32 and 42 to include a wide range of compression levels. The compression mode was set to a constant QP to maintain a consistent level of quantization errors. The depth maps were compressed with the H.264/AVC standard, using the reference software JM (version 15.1) [9].

For rendering purposes, the original depth maps and their compressed versions were utilised within the ViSBD software [5]. The rendered views were generated at the locations were original colour views existed, within the available MVD setup. All generated views were synthesised in an extrapolation rendering scenario, i.e. the rendered views were synthesised using one original colour view and its corresponding depth maps (original and compressed versions of the depth map). The mask replacement approach proposed within this paper was then applied to the resulting rendered views.

The views generated (non-masked views), together with their equivalent masked versions, were utilised in the video subjective assessment. Single frames, were extracted from both the non-masked views and masked versions of the rendered views, for the purpose of the image subjective assessment.

3.2 Subjective assessment procedure

Once all the dataset preparation steps were implemented, subjective assessment sessions of all video and image datasets were carried out. All image and video assessments followed the same procedure. The test environment and setup confirmed to the laboratory general viewing conditions as detailed in the ITU-R BT.500-13 [10]. Eighteen female/male observers, with ages ranging between 21 and 41 years old, took part in all of the subjective assessments.

A total of four subjective assessments were carried out: two for video (masked and non-masked assessments), similarly two more sessions were carried out for the image assessment. The Single Stimulus (SS) method was used for all the test sessions, with the original sequences at the rendered locations being used as hidden references. A continuous quality grading scale was used to record the opinions of the observers.

A 46” JVC Full HD LCD was employed for the purpose of displaying the test sequences. Test sessions consisted of 35 video/image sequences each. The observers were introduced to the test environment, voting interface, grading scale and were presented with a sample of the test sequences (training sequences). Test sequences in all assessments were displayed in random order to each observer.

3.3 Objective measurement

As with subjective evaluation both the masked and non-masked rendered views were objectively evaluated, against the original views present at the targeted rendering location. The obtained measurement results are analysed together with the subjective results in order to obtain an indication of the performance of the proposed mask replacement approach.

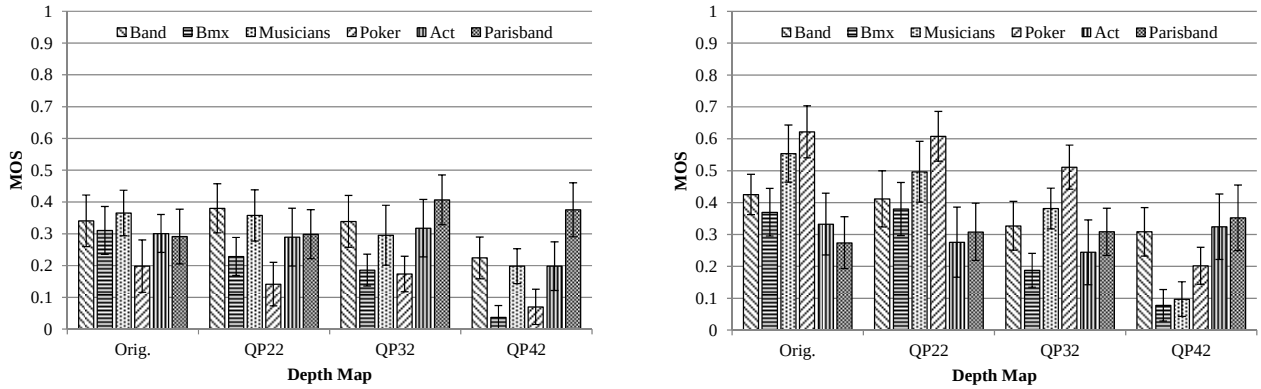
For the purpose of objective evaluation, four of the most extensively used image/video quality measurement techniques are utilised within this paper. These techniques are namely: Peak Signal to Noise Ratio (PSNR), Peak Signal to Perceptible Noise Ratio (PSPNR) [11], Structural Similarity Index (SSIM) [12] and Video Quality Metric (VQM) [13].

4. RESULTS AND DISCUSSION

In this section, the results of the subjective assessments are presented, together with the observations drawn from the results. Correlation with objective measurements is used to provide further insight into the observations obtained from the subjective results.

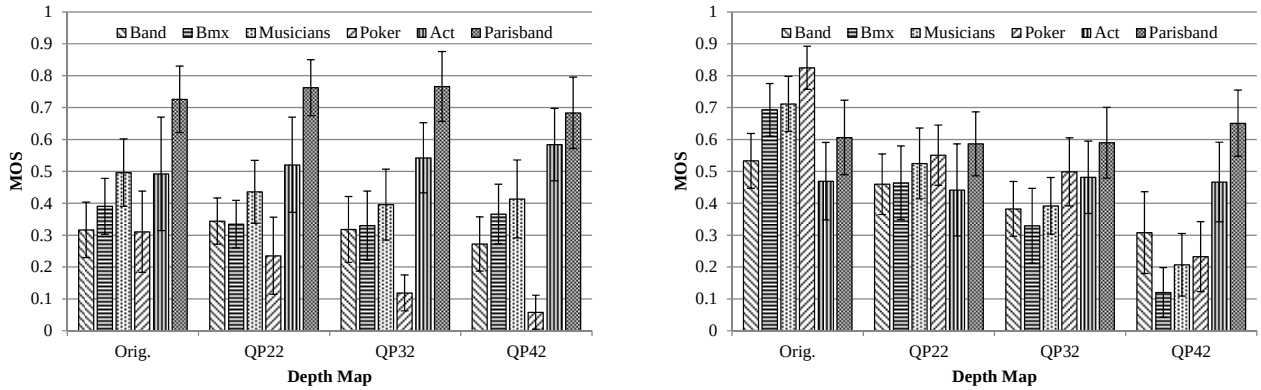
The first stage of the subjective results analysis includes the calculation of the Mean Opinion Scores (MOS), for each of the test sequences presented in all of the video and image assessment sessions. Standard deviation and a 95% confidence interval were also calculated. Observer screening was applied to the results to detect the outliers and no observers were rejected. All the results analysis processes were carried out in accordance with the calculations outlined in the ITU-R BT.500-13 [10]

Figure 2 shows the MOS results of all subjective assessments, together with the 95% confidence interval for each test sequence. The depth maps utilised in rendering were used to identify the test sequence scores (i.e. Orig. represents the sequence rendered using the original depth, QP22 is rendered using the compressed version of the depth map with the QP value set to 22, and so on).



(a) Non-masked video subjective results.

(b) Masked video subjective results.



(c) Non-masked image subjective results.

(d) Masked image subjective results.

Figure 2. Subjective assessment results.

The non-masked video/image subjective assessment results, shown in Figures 2(a) and 2(c), demonstrate no clear pattern in the MOS values for either the videos and images rendered using the different versions of the depth maps. This confirms the fact that the disocclusion regions within a rendered view - especially views rendered in the extrapolation scenario - have a more prominent visual impact on the overall quality of the rendered view, than other distortions within that view, which are caused by depth map inaccuracies.

On the other hand, in the masked subjective assessment results, shown in Figures 2(b) and 2(d), a clearer pattern in the subjective scores can be noticed. The videos and images rendered using the highly compressed depth maps, (i.e. depth maps compressed by setting a high QP value: namely QP32 and QP42), have been generally penalised by the observers. These views received a lower MOS value when compared to the views rendered using the original depth map and the depth map compressed using a QP value of 22. Although this pattern is generally observed from both the video and image masked results, it is more evident when *Band, BMX, Musicians and Poker* sequences are specifically considered.

The observation made from both the masked and non-masked assessment results, supports the fact that the proposed mask replacement approach eliminates the prominent effect of the disocclusion regions within the rendered views. Meanwhile the proposed approach application does not affect the presence of distortions and artefacts, caused by depth map inaccuracies.

The observations made from the subjective assessment results, shown in Figure 2, are further ratified by obtaining the objective measurements for the rendered views. The objective measurements employed were PSNR, PSPNR, SSIM and VQM. At this stage objective measurement was applied conventionally, i.e. without the application of the proposed mask replacement approach. The correlation between the non-masked objective measurements and the subjective assessment results is shown in Table I.

Table I, confirms the observations made from the subjective results, this is evident from the higher correlation coefficients between the masked image/video subjective results and all the non-masked objective measurements (obtained with no application of the proposed mask replacement approach). PSPNR and VQM measurements are Not Available (NA) for the image case, as these objective metrics are video quality metrics only.

Table I. Correlation between the objective and subjective results

	Non-mask video non-mask metric	**Mask video non-mask metric**	**Mask video mask metric**	**Non-Mask image non-mask metric**	**Mask image non-mask metric**	**Mask image mask metric**
Correlation Coefficient (CC)						
PSNR	0.4821	0.5654	0.5564	0.7227	0.7536	0.776
PSPNR	0.3071	0.4599	0.7872	NA	NA	NA
SSIM	0.2476	0.381	0.381	0.5021	0.5776	0.5353
VQM	0.3646	0.4943	0.6835	NA	NA	NA
Root Mean Square Error (RMSE)						
PSNR	0.084	0.1149	0.1147	0.1204	0.1134	0.1025
PSPNR	0.0913	0.1225	0.0851	NA	NA	NA
SSIM	0.0929	0.12757	0.1276	0.1491	0.1405	0.1373
VQM	0.0893	0.12	0.1008	NA	NA	NA

Furthermore, application of the proposed mask replacement approach to the utilised objective measurements provides interesting results. The correlation analysis between the masked objective measurements and the masked subjective results show a significant increase in the correlation coefficient between the masked video results and both the PSPNR and VQM masked metrics. This can be seen in Table I. (cells highlighted in grey). It is worth noting that PSPNR and VQM are objective measurements that account for the temporal quality of a processed video sequence within their resulting quality score.

This final observation indicates that the impact of the disocclusion regions on the temporal quality of the rendered views is greater than its effect on the spatial quality. This is evident in terms of both the subjective assessment scores and the objective evaluation.

5. CONCLUSION

This paper has presented a mask replacement approach that is targeted at addressing the visual impact, disocclusion regions (present within rendered views) have on the subjective assessment scores. The approach is based on the idea of replacing the occluded regions within an extrapolated view, with colour information from an original colour view, which is present at the location of the extrapolated view. This approach has enabled the subjective assessment scores of extrapolated rendered views, to provide an improved representation of the quality of depth maps used in the rendering process. The proposed approach has been utilised in both subjective assessment and objective evaluation of the rendered colour views.

The experimental results have provided a clear indication that the proposed mask replacement approach can deliver improved correlation results. Thus, this proposed approach should be taken into consideration as a valid tool with respect to subjectively evaluating the quality of a depth map. This approach will facilitate a more accurate examination of the concept of depth map quality.

6. ACKNOWLEDGMENT

This work was supported by the ROMEO project (grant number: 287896), which was funded by the EC FP7 ICT collaborative research programme.

7. REFERENCES

[1] E. Bosc, R. Pepion, P. Le Callet, M. Koppel, P. Ndjiki-Nya, M. Pressigout, L. Morin, "Towards a New Quality Metric for 3-D Synthesized View Assessment," Selected Topics in Signal Processing, IEEE Journal of , vol.5, no.7, pp.1332,1343, Nov.2011 doi:10.1109/JSTSP.2011.2166245.

[2] L. Po; S. Zhang; X. Xu; Y. Zhu, "A new multidirectional extrapolation hole-filling method for Depth-Image-Based Rendering," Image Processing (ICIP), 2011 18th IEEE International Conference on, pp.2589, 2592, 11-14 Sept.2011 doi:10.1109/ICIP.2011.6116194.

[3] M. Koppel, X. Wang, D. Doshkov, T. Wiegand, P. Ndjiki-Nya, "Consistent spatio-temporal filling of disocclusions in the multiview-video-plus-depth format," Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on, pp.25,30, 17-19 Sept. 2012. doi: 10.1109/MMSP.2012.6343410.

[4] L. Wang; J. Liu; J. Sun; Y. Ren; W. Liu; Y. Gao, "Virtual view synthesis without preprocessing depth image for depth image based rendering," 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2011, pp.1,4, 16-18 May 2011 doi: 10.1109/3DTV.2011.5877155

[5] MPEG. “View Synthesis Based on Disparity/Depth Software” (2008). [Hosted on MPEG server]. Available: http://wg11.sc29.org/svn/repos/MPEG4/test/trunk/3D/viewsynthesis/ViSD

[6] V. De Silva, C. Kim, N. Haddad, E. Ekmekcioglu, S. Dogan, A. Kondoz, I. Politis, A. Kordelas, T. Dagiuklas, M. Weitnauer, and C. Hartmann, "An End-to-End QoE Measurement Framework for Immersive 3D Media Delivery Systems", Proceedings of the 1st ROMEO Workshop, Athens, Greece, 9 July 2012.

[7] MUSCADE deliverable D2.2.2. “3D Capture and post-production development - Phase1”. (2012). [online]. Available: http://www.muscade.eu/deliverables/D2.2.2.pdf.

[8] ROMEO deliverable D3.1. “Report on 3D media capture and content preparation”. (2012). [online]. Available: http://www.ict-romoe.eu/pdf/D3.1.pdf.

[9] H.264/MPEG-4 AVC Reference Software Manual. Dolby Laboratories Inc., Fraunhofer-Institute HHI, Microsoft Corporation. (2009) [Online]. Available: http://iphome.hhi.de/suehring/tml.

[10] International Telecommunication Union/ITU Radio communication Sector, “Methodology for the subjective assessment of the quality of television pictures”, ITU-R BT.500-13, 2012.

[11] Y. Zhao, L. Yu, "PSPNR Tool 2.1, R. software JVT-X208, in ISO/IEC JTC1/SC29/WG11, MPEG09/N10879, Kyoto 2009.

[12] J. Klaue, B. Rathke, and A. Wolisz, "EvalVid - A Framework for Video Transmission and Quality Evaluation", 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation, pp. 255-272, Urbana, Illinois, USA, September 2003.

[13] American National Standards Institute, “American National Standard for Telecommunications – Digital transport of one-way video signals – Parameters for objective performance assessment,” ANSI T1.801.03 – 2003.