2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)

Subjective Assessment of Stereoscopic Image Quality: The Impact of Visually Lossless Compression

Sanjida Sharmin Mohona
Dept. of Electrical Engineering and Computer Science
York University
Toronto, Canada
sanjida.sharmin23@gmail.com

Domenic Au
Department of Biology
York University
Toronto, Canada
domau@my.yorku.ca

Onoise Gerald Kio
Dept. of Electrical Engineering and Computer Science
York University
Toronto, Canada
ogkio@cse.yorku.ca

Richard Robinson
Dept. of Electrical Engineering and Computer Science
York University
Toronto, Canada
rir@my.yorku.ca

Yuqian Hou
Centre for Vision Research
York University
Toronto, Canada
yhou@yorku.ca

Laurie M. Wilcox
Department of Psychology
York University
Toronto, Canada
lwilcox@yorku.ca

Robert S. Allison
Dept. of Electrical Engineering and Computer Science
York University
Toronto, Canada
allison@eecs.yorku.ca

***Abstract*—In stereoscopic displays different images are presented separately to the left and right eyes. This requirement may increase the bandwidth demand as well as increase the occurrence of visible compression-related artefacts. Here we report the results of a large-scale subjective assessment of high dynamic range (HDR) stereoscopic image compression. The ISO/IEC 29170-2 flicker paradigm was adapted for stereoscopic images and used to evaluate two VESA (Video Electronics Standards Association) image compression codecs: DSC 1.2a and VDCM 1.2.2. We compared the performance on stereoscopic images versus 2D images for both codecs.**

Keywords— Display stream compression (DSC, VDC-M), stereoscopic display, subjective assessment, visually lossless.

I. INTRODUCTION

The technology and market demand for high dynamic range (HDR) images, wide color gamut, and stereoscopic 3D (S3D) augmented and virtual reality displays are growing rapidly. These advances have increased the bandwidth demand across the display link, which has intensified the need for low-latency, high-quality image compression. Industry has responded with new standards, for example two state-of-the-art image compression codecs, DSC (Display Stream Compression) and VDC-M (VESA Display Compression-M), have been developed by the Video Electronic Standard Association (VESA) to meet these requirements. Both codecs have undergone rigorous objective and subjective assessment for 2D standard dynamic range (SDR) [1] and high dynamic range (HDR) [2] images at target compression levels using the ISO/IEC 29170-2 [3] protocol. However, the performance of these codecs for stereoscopic 3D (S3D) images cannot be necessarily predicted from the codecs' performance for 2D images; S3D images are processed differently than 2D images by the human visual system [4], [5]. Under stereoscopic viewing conditions (as in virtual reality) each eye sees a different image and depth variation in the scene introduces differences between them. Subjective assessment of S3D SDR images has been performed for both DSC [6] and for VDC-M [7] using a novel implementation of the ISO/IEC flicker paradigm. But these studies have been limited in scope and to date no such evaluation has been performed with HDR S3D content.

In this paper we present the results of two experiments, each with a large number of naïve subjects, in which we conducted subjective assessments of DSC 1.2a and VDC-M 1.2.2. In both experiments we tested S3D HDR images at two compression levels for each codec to evaluate the relative visibility of compression artefacts under 2D and S3D viewing conditions. Computer-generated S3D HDR images were generated to span a wide range of image content and expected to challenge the codecs. The test scenes were designed to represent typical use cases and to target important qualitative classes of interocular image differences encountered in stereoscopic viewing.

II. BACKGROUND

In a stereoscopic 3D display each eye is given a separate image and the human visual system processes differences or disparities between the left eye and right eye images to perceive depth or 3D structure. As the two images in a stereoscopic pair are not identical, coding artefacts can differ between the images or can be common [8]. The visibility of these artefacts cannot be predicted by 2D objective measurements as these do not account for the binocular characteristics of the HVS. When there are distortions in the shape, size or position of portions of one eye's image relative to that of the other image, stereoscopic depth distortion can arise [5], [9]. This kind of depth distortion appears within an object or scene and may in turn distort perceived shape. Generally, global depth distortions are well-tolerated by users [9] but compression may produce local distortion artefacts which are generally more visible. Furthermore, when a feature is introduced or removed in only one eye's image, spurious monocular features arise. In natural scenes monocular features are most common at depth edges where there is occlusion in one view but visibility in the other [10]. Behavioural studies have shown that when such features are consistent with the viewing geometry, depth change at the edge is perceived appropriately [11], [12]. However, the monocular features introduced through compression can be spurious and

Financial support from the Ontario Centres of Excellence through the VIP II program and from VESA through a contract are gratefully acknowledged.

978-1-7281-5965-2/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: York University. Downloaded on June 23,2022 at 16:31:48 UTC from IEEE Xplore. Restrictions apply.

inconsistent with the 3D layout of the scene. Other interocular differences can arise from view dependent material properties. Highlights on shiny surfaces can cause intensity variation between the two eyes that appears as lustre. If these regions are affected by compression artefacts, then a shiny surface may appear as dull or a matte surface may appear as shiny. Such artefacts may be more common in HDR images because in these images surface material properties can be represented with higher fidelity. Another visual phenomenon that may impact (and in this case reduce) the visibility of compression artefacts in S3D is dichoptic masking. This occurs when two images of a similar pattern are shown to each eye and the detection of the test stimulus in the one eye is reduced by the 'mask' stimulus in another eye. When the images are similar, masking is most pronounced. Dichoptic masking should make stereoscopic viewing more robust to compression artefacts than 2D images.

In sum, when considering S3D vs 2D the potential for visible image artefacts increases substantially but dichoptic masking may make them less objectionable. Subjective testing is essential understand how these factors interact to determine image quality in S3D.

III. Method

A. Stimuli

A set of test scenes was designed and rendered to sample images that targeted the S3D issues raised in Section II. Rendered images were used to conform with the target VR/AR application domain but also to allow for flexible and precise control of image parameters. Custom-designed scenes were combined with scenes and objects purchased or freely available on online content marketplaces to create the desired scenarios in 3D graphics applications. From these scenes, 132 S3D reference images were rendered using Blender™, Unity™, and Unreal Engine™. Camera positioning and rendering were aimed at generating specific features in the images including (1) monocular zones, (2) overlays or heads-up displays (HUD), and (3) fine textured surfaces.

Within these broad categories care was taken to include specific details such as (1) specularities and highlights, occlusion of texture/objects in monocular regions; (2) overlays/HUDs which are conformal, separated in depth from the scene, or monocular, and (3) separation in depth between finely textured surfaces, depth variation within a textured object, and natural 3D textures such as foliage. In the evaluation of candidate images, we checked the stereoscopy and emphasized high-quality and comfortable S3D image pairs and ensured that double vision (diplopia) or edge violations were not introduced by the S3D image generation and display. All images were compressed and then objective analysis with PSNR (peak signal-to-noise ratio) and S-CIELAB were performed to identify challenging images for the codec. Typical images are compressed with minimal distortion as expected for a visually lossless codec. Even though our scenes were designed to be difficult to compress this minimal distortion was true for most of the test images as well. Thus, compressed images with very high PSNR were excluded as uninteresting due to lack of artifact (confirmed by visual inspection). Further image selection was based on visual assessment of the presence of image artefacts by four expert observers. Following this process, ten S3D HDR images that challenged the codecs were selected from the initial set of 132 images. The rendered reference images were 4k resolution (3840 X 2160 pixel), BT 2020 colour space, PQ (SMPTE 2084) transfer function, HDR images (4:4:4 pixel sampling, 10 bits per color channel and 30 bits per pixel (bpp)). Each image was compressed using DSC 1.2a and VDC-M 1.2.2 at the native compression level (i.e. 8 bpp for DSC and 6 bpp for VDC-M) and at the estimated breakpoint (at limit of visually lossless compression) or just below the breakpoint if breakpoint and native compression level were same. The full-frame images were compressed with codec settings of 2 slices per line and a slice height of 108 pixels. Both the full size (3840 X 2160) original (reference) and compressed images were cropped to a 1200 X 1000 pixels region. The crop regions were selected in a pilot study where full size original images were interleaved with full-size compressed images at 5 Hz. These were viewed by experts and the regions containing noticeable flicker (most visible compression artifacts) were selected as regions of interest and cropped. The purpose of cropping was to direct the participant's attention to these regions of interest [3].

B. Observers and Apparatus

We screened all observers for stereoacuity (at least 40 arc seconds for inclusion), visual acuity (at least 20/25 for inclusion) and color vision using Snellen Chart, Randot stereo test (2015 Stereo Optical Company, Inc.) and Ishihara tests, respectively. Based on the visual screening, nine subjects were excluded (one for visual acuity, seven for stereo acuity and one for color vision). After the visual screening we tested 60 (28 Female, 32 Male; median age 25 years (between 18 to 35 years)) subjects for the experiment, all were from the York University Community.

Two test stations (65"W X 97"H X 98"L) used in the study. All sides of each booth were covered by floor length dark gray curtains and the study was conducted in a dark room. Each testing station contained a fully-calibrated mirror stereoscope with dual ASUS PA32UCX HDR monitors (70.8 X 39.8 cm @60 Hz, screen resolution: 3840 X 2160 pixels, color: P3 (BT2020 container), 10-bit PQ, peak luminance 1000 cd/m^2) viewed through fully-silvered mirrors placed at ±45º to the observer's face, which was used to view the S3D test images. The background luminance for both monitors was 0.01 cd/m^2. The host computers were Intel i7-9700k, NVIDIA GeForce RTX 2080 GPUs, running Windows 10, 64-bit operating system. The distance to the surface of the displays (viewing distance) was 63 cm resulting in a resolution of 60 pixel per degree (ppd) at the centre of the display. A chinrest was used to stabilize the subject's head and position the eyes in the stereoscope. Custom C++ code based on DirectXTK HDR was used to display the stimuli and record subject's response. Subjects used a Xbox controller gamepad to provide their responses and Philips SHP1900 over-ear head headphones were used to block environmental noise and present feedback for incorrect responses (500 Hz, 0.1 s tone).

C. Procedure

We tested each image at two compression levels (native compression level and breakpoint) for each codec. In each experiment there were three viewing conditions: 3D, 2D left eye only, 2D right eye only. In the 2D conditions both eyes viewed the same image in the stereoscope. We divided the participants into three groups and in each of the three groups we tested eight conditions (i.e., four images at each of two

Authorized licensed use limited to: York University. Downloaded on June 23,2022 at 16:31:48 UTC from IEEE Xplore. Restrictions apply.

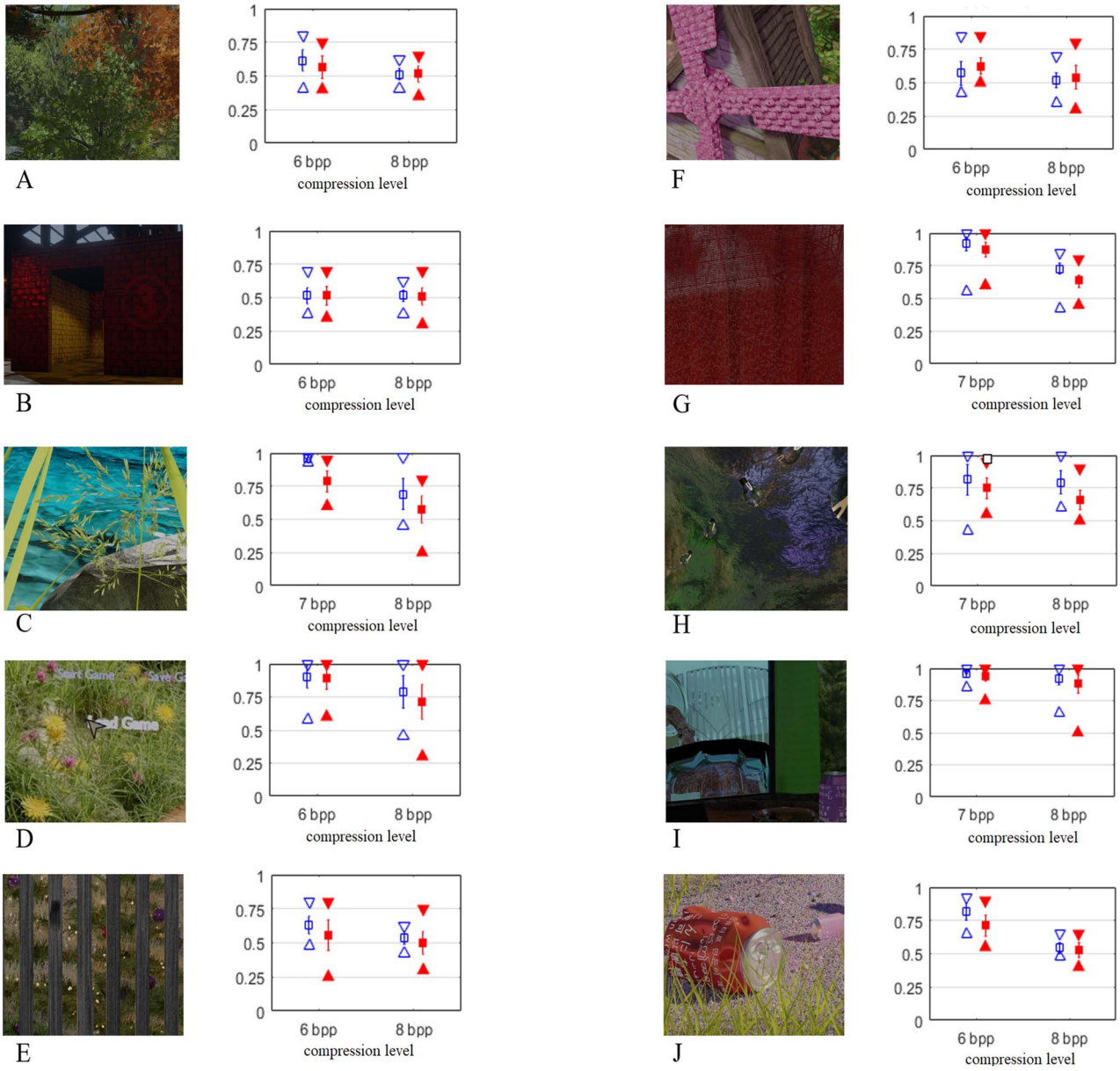

Fig. 1. Flicker detection rates for 10 test images in experiment 1 (A. BigTree2_30, B. HangarHUD, C. beach, D. fox, E. picketfence, F. DiscoWindMill_30, G. bathroom, H. cave, I. monitor, and J. rocks) for DSCv1.2a. The x-axis represents compression level, y-axis represents average proportion of correct detection, red color represents S3D viewing and blue color represents 2D viewing. Square symbols represent the proportion correct averaged across all observers. The error bars represent ± 1 standard deviation, and triangles indicate the best and worst performance.

compression levels). Each condition was repeated 20 times for each participant and we also included two catch trial images (compressed with noticeable JPEG 2000 compression) each of which was also repeated 20 times. Thus, there were 480 test trials and 40 catch trials per participant in each group, for a total of 520 trials. These 520 trials were divided into ten blocks of 52 trials and each trial was displayed for 8 s. Each observer participated in two sessions: in the first session we conducted visual screening and tested 6 blocks, in the second session 4 blocks were tested. There was at least a one-hour break between sessions.

Observers performed a two-alternative forced-choice task, which was a modified version of the ISO/IEC 29170-2 (Annex B) protocol [3], following the approach used by [6] and [7]. Test-image and reference-image sequences were displayed side by side and subjects were asked to indicate which image was flickering. In test-image sequences, the compressed version of the image was interleaved with the uncompressed version and in the reference-image sequences the uncompressed version was interleaved with itself, in both cases the interleave frequency was 5 Hz (100 ms for each phase of the alternation). Subjects provided their response either within the 8 s viewing period or following it. The feedback tone was played if viewers made an incorrect response.

IV. Results

Fig. 1 and Fig. 2 show the average proportion correct for each condition with DSC1.2a and VDC-M 1.2.2 compression, respectively, plotted after the ISO/IEC 29170-2 recommendation. For each condition, the square symbol represents the mean proportion correct score, the error bar represents ± 1 standard deviation, and the triangles denote the range of scores. The best performing observer (for whom the

Authorized licensed use limited to: York University. Downloaded on June 23,2022 at 16:31:48 UTC from IEEE Xplore. Restrictions apply.

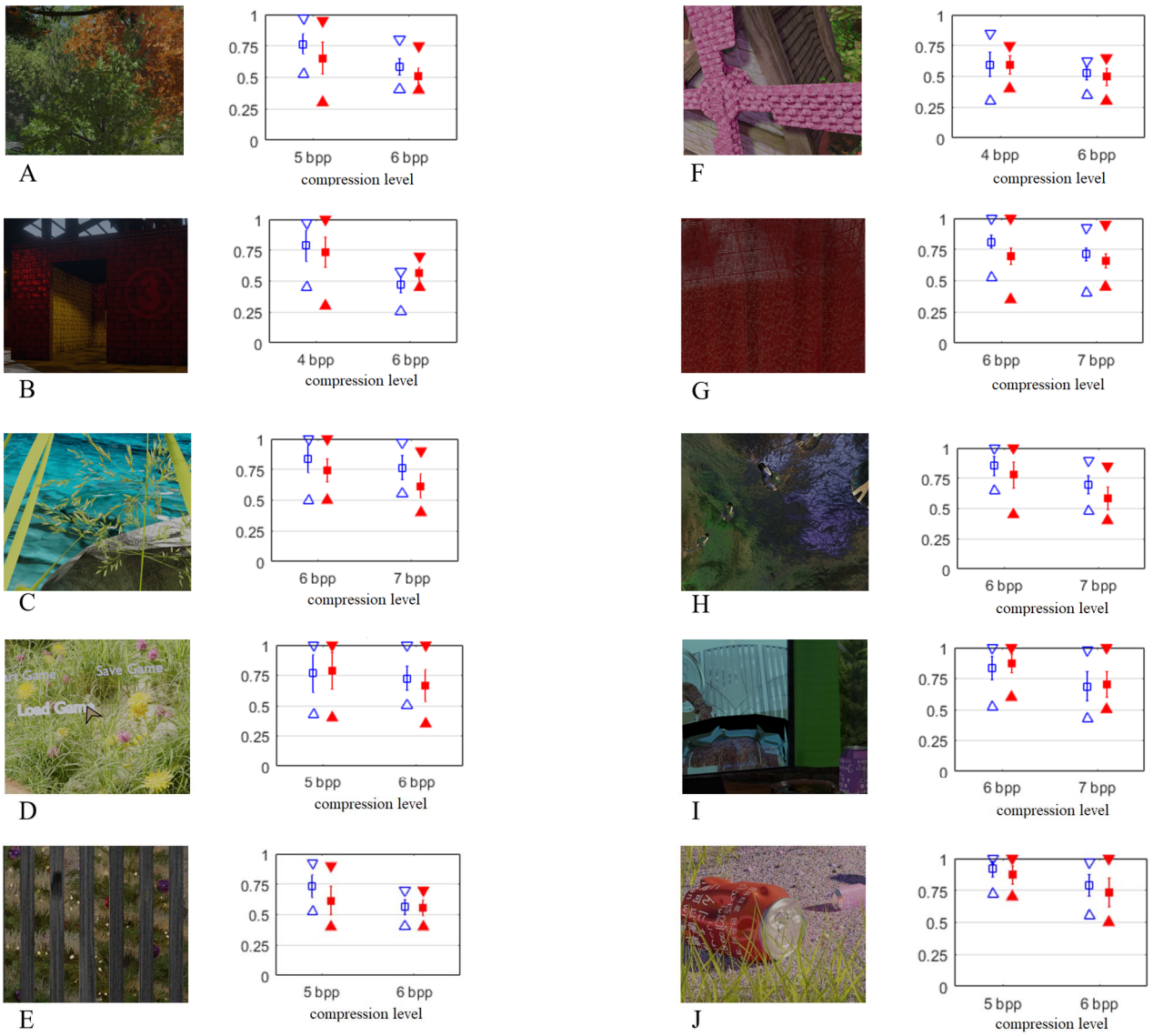

Fig. 2. Flicker detection rates for 10 test images in experiment 2 (A. BigTree2_30, B. HangarHUD, C. beach, D. fox, E. picketfence, F. DiscoWindMill_30, G. bathroom, H. cave, I. monitor, and J. rocks) for VDC-M V1.2.2. The x-axis represents compression level, y-axis represents average proportion of correct detection, red color represents S3D viewing and blue color represents 2D viewing. Square symbols represent the proportion correct averaged across all observers. The error bars represent ± 1 standard deviation, and triangles indicate the best and worst performance.

proportion correct score is highest) is indicated by a downwards triangle and worst performing observer (for whom the proportion correct score is lowest) by an upwards triangle. We hypothesized that the perception of artefacts would differ between 2D and S3D viewing conditions for each of the codecs. We fitted the data using a Generalized Linear Mixed Model (GLMM) [13] following the approach described by Cutone et al. [14]. As data for catch trials are unrelated to the experimental hypothesis and may cause convergence failure, we excluded these data from the analysis. The following model was used to assess the overall effects of depth on artefact visibility:

$$\text{Correct} \sim \text{depth} + (1|\text{Subject}) + (\text{depth}\,|\,\text{Condition}) \qquad (1)$$

For the above formula, depth is considered as fixed effect and Condition (combination of image and compression level) & Subject ID were considered as random factors. The responses were dichotomous (correct or not), so the error distribution of the response variable was modeled as binomial. The ANOVA for the above formula is in Table I. The analysis showed that the overall effect of depth on artefact detection was significant for both DSC 1.2a and VDC-M 1.2.2. A series of planned comparisons between the proportion correct for the 2D and S3D viewing for each condition (i.e. for each image-codec compression level) were performed. This analysis was based on tests of linear contrasts estimating the difference between 2D and S3D viewing for each condition. We used the lsmeans package in R to obtain the least square mean predictions. For this testing we applied False Detection Rate (FDR) p-value correction at a significance level of 0.05.

Fig. 3 and Fig. 4 show the comparison between 2D and S3D viewing for DSC 1.2a and VDC-M 1.2.2 respectively. For seven conditions under DSC compression (indicated by red rectangles in Fig.3), there were significant differences in artefact detection rate between 2D and S3D viewing. In all these cases artefacts were less perceptible in S3D compared to 2D viewing. Similarly, Fig. 4 shows that for nine conditions (indicated by red and blue rectangles), there was a significant difference in the visibility of 2D and S3D artefacts for VDC-M 1.2.2. Among these nine conditions artefacts were again less perceptible in S3D for all conditions except one (indicated by blue overlay in Fig. 4).

Authorized licensed use limited to: York University. Downloaded on June 23,2022 at 16:31:48 UTC from IEEE Xplore. Restrictions apply.

Fig. 3. Average proportion correct for each condition for the 2D vs 3D viewing in DSC 1.2a (data for 2D left eye and 2D right eye are collapsed). Error bars represent ±1 Standard Error (SE).

Fig. 4. Average proportion correct for each condition for the 2D vs 3D viewing in VDC-M 1.2.2 (data for 2D left eye and 2D right eye are collapsed). Error bars represent ±1 SE.

Authorized licensed use limited to: York University. Downloaded on June 23,2022 at 16:31:48 UTC from IEEE Xplore. Restrictions apply.

TABLE I. OVERALL EFFECT OF DEPTH ON ARTEFACT DETECTION FOR DSC1.2A AND VDCM 1.2.2. SIGNIFICANCE: '***' P<0.001 '**' P<0.01.

Codec		$\chi^2(1)$	Pr(>Chisq)
DSC1.2a	(Intercept)	24.695	6.717e-07 ***
	depth	15.978	6.409e-05 ***
VDCM1.2.2	(Intercept)	39.116	3.994e-10 ***
	depth	16.567	4.695e-05 ***

(a) (b)

Fig. 5. HangurHUD: (a) left eye image, (b) right eye image

V. DISCUSSION

These experiments represent the first assessment of the effects of display stream compression on the perception of artefacts in S3D HDR imagery. We found that at the native compression level of DSC1.2a (i.e., at 8bpp) all images except monitor showed visually lossless performance in the S3D viewing condition. For five images (bathroom, beach6, cave, monitor, rocks) in the S3D viewing condition, VDC-M 1.2.2 did not show visually lossless performance at native compression level (i.e., at 6bpp).

The images which did not show visually lossless performance have either specularities and/or fine texture. The main objective was to determine the relative visibility of 2D and S3D artifacts. The results are consistent with our previous observation [7] that there is significant perceptual silencing under S3D viewing (compared to the same content viewed in 2D). There are some common features among the images which showed significant perceptual silencing in S3D viewing, e.g., specularities/highlight (beach6, cave, rocks), fine texture (bathroom, rocks), foliages (beach, fox, rocks, picketfence). The single exception to this result is an image (shown in Fig. 5) which contains significant differences in monocular content (an additional luminous green reticle in one image), which suggests that displays with significant differences between the left and right displays (here a monocular HUD) may be more susceptible to compression artefacts. As seen in Fig. 5, the monocular content in right eye was emitting light (glowing). Interestingly, there was no significant difference in 2D performance between the left and right images for this case, despite the large differences in monocular appearance. This suggests that the binocular combination of the compressed images makes the compression distortions more apparent rather than these distortions being particularly salient in either of the images. It is also important to note that the codecs performed well for this image and exhibited visually lossless performance under both 2D and S3D viewing conditions at native compression level.

The artefact regions identified in the test images varied both across and within categories. Thus, it was difficult to predict the artefacts based on the category for which they were generated. Instead, it appears that artefacts are image specific and likely determined by more low-level image properties and how these are processed binocularly. Insight into this issue will be important for the use of visually lossless codecs in stereoscopic display systems. Of particular interest is virtual reality where artefacts may be exacerbated by additional factors such as optical distortion, color conversion, limited display resolution and observer motion.

REFERENCES

[1] R. S. Allison *et al.*, "75-2: Invited Paper: Large Scale Subjective Evaluation of Display Stream Compression," *SID Symposium Digest of Technical Papers*, vol. 48, no. 1, pp. 1101–1104, May 2017, doi: 10.1002/sdtp.11838.

[2] A. Sudhama *et al.*, "85-1: Visually Lossless Compression of High Dynamic Range Images: A Large-Scale Evaluation," *SID Symposium Digest of Technical Papers*, vol. 49, no. 1, pp. 1151–1154, doi: 10.1002/sdtp.12106.

[3] ISO/IEC 29170-2, "ISO/IEC 29170-2:2015 - Information technology -- Advanced image coding and evaluation -- Part 2: Evaluation procedure for nearly lossless coding," 2015. [Online].

[4] D. Kane, P. Guan, and M. S. Banks, "The Limits of Human Stereopsis in Space and Time," *J. Neurosci.*, vol. 34, no. 4, pp. 1397–1408, Jan. 2014, doi: 10.1523/JNEUROSCI.1652-13.2014.

[5] I. P. Howard, and B. J. Rogers, *Binocular Vision and Stereopsis*. Oxford University Press, 1995.

[6] S. S. Mohona, L. M. Wilcox, and R. S. Allison, "Subjective Assessment of Display Stream Compression for Stereoscopic Imagery." *IEEE transaction on image processing(submitted)*, January 2020.

[7] D. Au *et al.*, "3-4: Stereoscopic Image Quality Assessment," *SID Symposium Digest of Technical Papers*, vol. 50, no. 1, pp. 13–16, Jun. 2019, doi: 10.1002/sdtp.12843.

[8] M. J. Chen, A. C. Bovik, and L. K. Cormack, "Study on distortion conspicuity in stereoscopically viewed 3D images," in *2011 IEEE 10th IVMSP Workshop: Perception and Visual Signal Analysis*, 2011, pp. 24–29, doi: 10.1109/IVMSPW.2011.5970349.

[9] R. S. Allison and L. M. Wilcox, "Perceptual Tolerance to Stereoscopic 3D Image Distortion," *ACM Trans. Appl. Percept.*, vol. 12, no. 3, pp. 10:1–10:20, Jul. 2015, doi: 10.1145/2770875.

[10] J. M. Harris and L. M. Wilcox, "The role of monocularly visible regions in depth and surface perception," *Vision Research*, vol. 49, no. 22, pp. 2666–2685, Nov. 2009, doi: 10.1016/j.visres.2009.06.021.

[11] B. Gillam and E. Borsting, "The Role of Monocular Regions in Stereoscopic Displays," *Perception*, vol. 17, no. 5, pp. 603–608, Oct. 1988, doi: 10.1068/p170603.

[12] K. Nakayama and S. Shimojo, "Da vinci stereopsis: Depth and subjective occluding contours from unpaired image points," *Vision Research*, vol. 30, no. 11, pp. 1811–1825, Jan. 1990, doi: 10.1016/0042-6989(90)90161-D.

[13] A. Agresti, "Random Effects: Generalized Linear Mixed Models," in *An Introduction to Categorical Data Analysis*, 2007.

[14] M. D. Cutone, M. Dalecki, J. Goel, L. M. Wilcox, and R. S. Allison, "P-31: A Statistical Paradigm for Assessment of Subjective Image Quality Results," *SID Symposium Digest of Technical Papers*, vol. 49, no. 1, pp. 1312–1314, 2018, doi: 10.1002/sdtp.12154.

Authorized licensed use limited to: York University. Downloaded on June 23,2022 at 16:31:48 UTC from IEEE Xplore. Restrictions apply.