Stereoscopic Image Quality Assessment

Domenic Au[1], Sanjida Sharmin Mohona[1], Matthew D. Cutone[1], Yuqian Hou[1], James Goel[2], Natan Jacobson[3], Robert S. Allison[1], Laurie M. Wilcox[1]

[1] Centre for Vision Research, York University, 4700 Keele St., Toronto, ON, Canada, M3J 1P3;
+1 416-736-5659; allison@cse.yorku.ca, lwilcox@yorku.ca

[2] Qualcomm Canada Inc., 100-105 Commerce Valley Drive West, Markham, Ontario, Canada, L3T 7W3;
+1 (905) 482-5765; jgoel@qti.qualcomm.com

[3] Qualcomm., 5775 Morehouse Dr, San Diego, CA 92121, USA;
+1 858-587-1121; natanj@qti.qualcomm.com

Abstract

Stereoscopic display technology provides immersive experiences in VR/AR/XR, but requires markedly higher bandwidth and is perceived differently than 2D content. Here we adapt the ISO/IEC 29170-2 flicker paradigm for subjective assessment of low impairment stereoscopic image compression. We compared the performance VESA VDC-M codec on stereoscopic images with 2D image performance.

Author Keywords

VESA Display Compression (VDC-M); specular highlights; visually lossless; subjective quality assessment; image compression; monocular occlusions

1. Objective and Background

The recent advances in and growing availability of stereoscopic 3D (S3D) augmented and virtual reality displays have increased the demand for efficient compression to support higher bandwidth across display links. Currently, however, there is no accepted protocol for evaluation of stereoscopic image quality. Recently developed lightweight compression algorithms, such as the VESA Display Compression-M Standard (VDC-M) [1,3], have been validated in subjective tests to be visually lossless under target compression levels. However, visually lossless performance with 2D displays does not necessarily predict similar S3D codec performance. For example, S3D stimuli are processed differently by the human visual system than 2D stimuli [4]. Further, S3D images that contain variations in depth also contain monocular differences between the left and right eye images [6]. These differences are particularly striking when specular highlights are present in either the left or right eye only. In this study, we conducted two subjective evaluations of VDC-M compression on images presented in 2D and S3D. In Experiment 1, we compared artefact detection in 2D and S3D viewing conditions using stereoscopic images of natural scenes; in Experiment 2, the same comparison was made but using computer-generated imagery with features that are visible to one eye only in a stereoscopic image pair. These experiments demonstrate that standardized forced choice procedures such as ISO 29170-2 can be adapted for evaluation of stereoscopic image quality. They also show that under the conditions tested here observers are relatively insensitive to compression artefacts in S3D content.

2. Methods:

2.1. Observers and Apparatus:

Subjects for both experiments were screened for color vision and acuity as specified in the ISO/IEC 29170-2 standard. They were also required to meet a criterion level of stereoacuity (40 arc seconds on the Randot™ stereo test [5]). Four participants were excluded based on this pre-screening and four participants were excluded due to poor performance on ‘catch’ trials. A total of 10 and 12 participated in Experiments 1 and 2 respectively. In both studies, images were presented on a mirror stereoscope using two HP Z Display Dream Color Monitors, 1920 x 1200 pixels, 52 x 32 cm, 60 Hz refresh rate. The computer was an Intel® Xeon® CPU E5-1620 v3 3.5 GHz, 16 GB memory, Windows 10, 64-bit operating system. The background luminance of both displays was 0.36 cd/m^2. Stimuli were presented using custom Matlab scripts at a viewing distance of 45 cm. The participant’s head position was stabilized using a chinrest.

2.2. Stimuli:

In both studies, image crops were tested in 2D and S3D. While image selection for traditional 2D subjective trials is challenging, it is even more so for S3D testing as one must also consider the quality of the stereophotography and the range of image disparity (depth). The S3D imagery used here contained uncrossed disparity (relative to the screen) with depth variation throughout the image. Candidate images were evaluated to ensure that both of the images in the stereopair were of high quality (e.g. in focus), did not introduce double vision (diplopia) or edge violations. These images were then compressed and evaluated using PSNR (peak signal to noise ratio) to determine whether the content would be challenging for the codec. In Experiment 1, 10 images were selected from 1000 candidates obtained from Flickr.com and the Middlebury image set. All images were stereoscopic photographs of natural scenes (Figure 2) [2]. In Experiment 2, 8 images were tested; 7 were created using Blender™ and 1 was selected from Flickr. For this image set, an emphasis was placed on rendering images that would produce monocular half-occlusions associated with specular highlights (Figure 1). This was done by manipulating the camera location, surface materials, and object surface textures. Each image was tested in S3D (stereoscopic image pairs) and 2D (left eye image presented to both eyes), with 20 trials per condition.

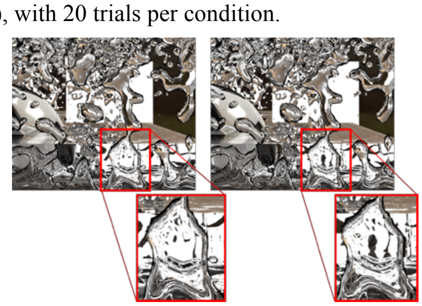

Figure 1. Fluid image rendered in Blender™ containing monocular differences between the left and right eye.

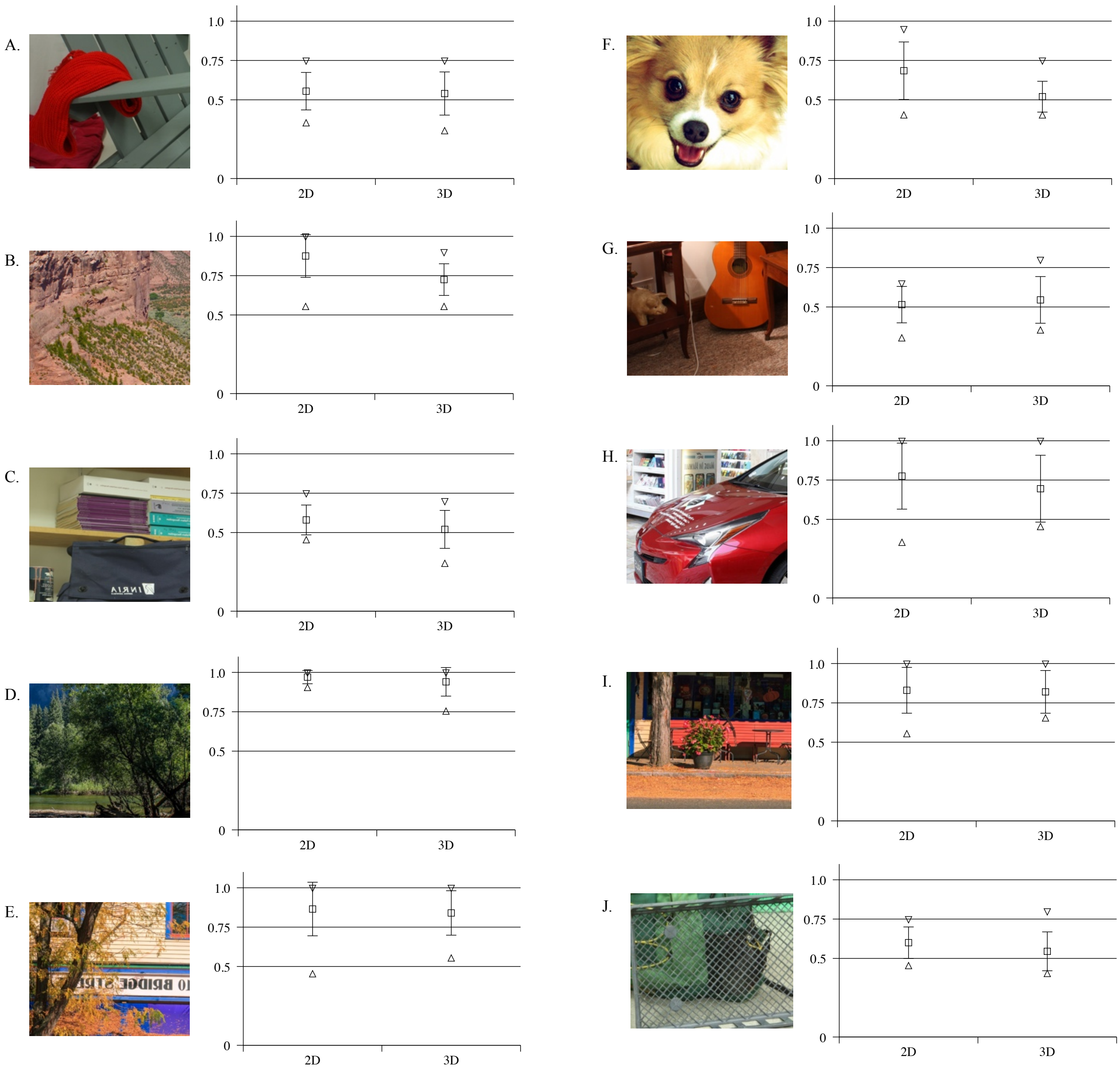

Figure 2. Flicker detection rates for 10 test images in experiment 1 (A. *Adirondack*, B. *RockMountain*, C. *Shelves*, D. *Tree*, E. *YellowTree*, F. *Aryaa*, G. *Piano*, H. *RedCar*, I. *Flower*, and J. *Backpack*) VDC-Mv1.1.0 at 4bpp. The 2D and S3D viewing conditions are indicated on the x-axis. Square symbols represent the proportion correct averaged across 10 observers. The error bars represent ± 1 standard deviation, and triangles indicate the best and worst performance.

VDC-M compression (v1.1.0 in Experiment 1, v1.2.0 in Experiment 2) was applied at 4bpp, 1 slice per line, slice height 108, 4:4:4 pixel sampling to the full frame images. A 600 x 500 pixel region was then cropped from both the original (reference) and compressed versions. As described in the 29170-2 protocol, images were cropped to focus participants' attention within a region of interest [5]. Crop regions were selected via preliminary testing with the full-scale image to identify regions that would best challenge the codec. The test and reference image sequences were presented side-by-side, centred on the midpoint of the display.

2.3. Procedure:

A two-alternative forced choice task, based on the ISO/IEC 29170-2 (Annex B) flicker protocol, was used in both experiments [7]. Observers were presented with two versions of the same image (the compressed target and the uncompressed reference) side-by-side, and each image alternated with the original (reference) image at a rate of 5Hz (for details see [7]). Because the images were presented using a mirror stereoscope, the image sequences were viewed by each eye separately; in the S3D conditions depth variation was evident within each image. As in the standard paradigm, observers were asked to indicate which image location (left side vs. right side) contained flicker. If the response was incorrect, auditory feedback was provided

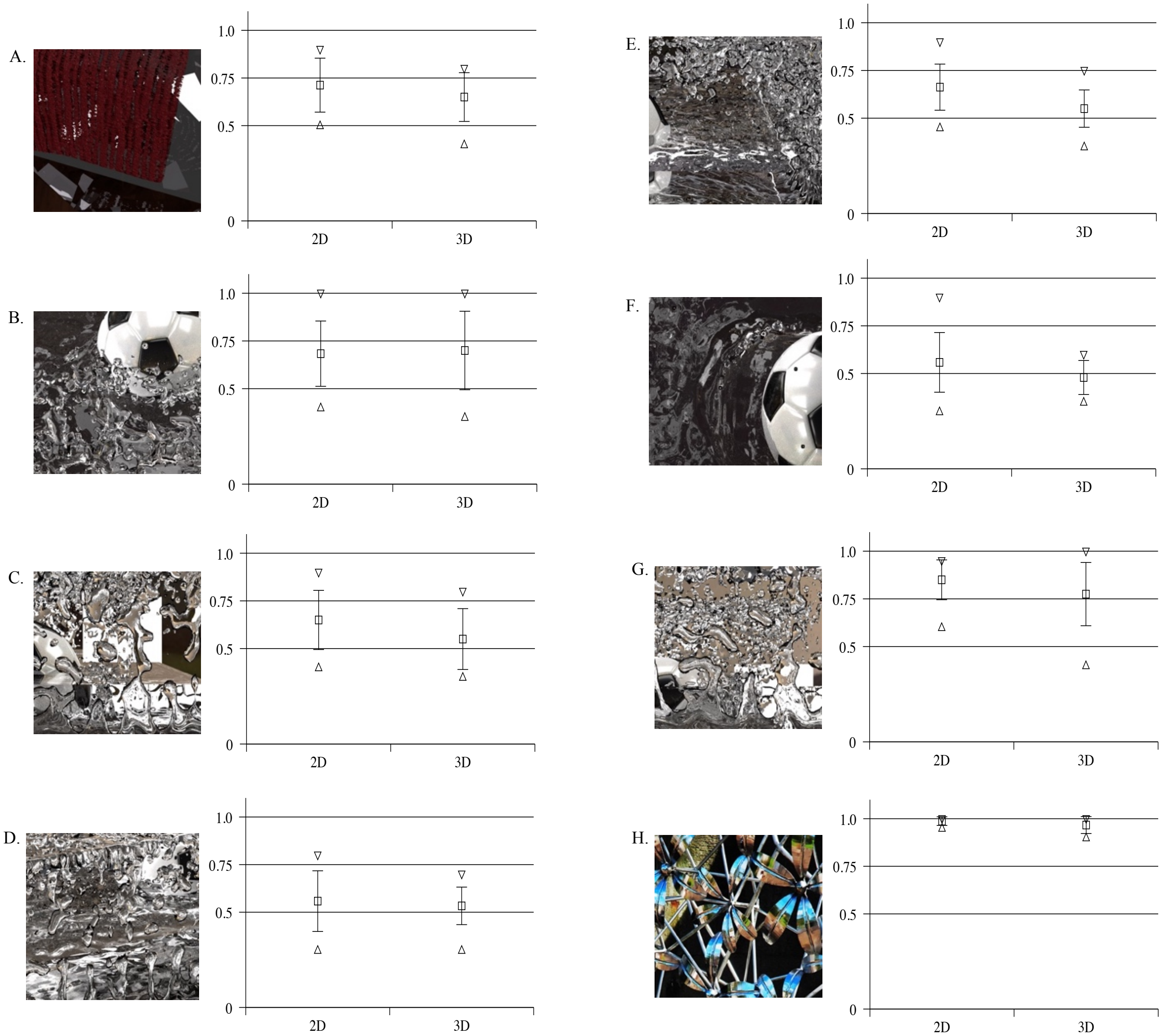

Figure 3. Proportion correct for flicker detection for 8 test images in experiment 2 (A. *Diamond*, B. *FluidA*, C. *FluidB*, D. *FluidC,* E. *FluidD,* F. *FluidE,* G. *FluidF,* and H. *Metalflowers*) VDC-Mv1.2.0 at 4bpp. The 2D and S3D viewing conditions are indicated on the x-axis. Square symbols represent the proportion correct averaged across 12 observers. The error bars represent ± 1 standard deviation, and triangles indicate the best and worst performance.

(500Hz, 0.1s tone). On each trial, the images were displayed for a maximum of 8 s, followed by a response screen. Each experiment was completed in two sessions.

3. Results:

Descriptive statistics were computed for all observers in each experiment and plotted using the ISO/IEC 29170-2 recommended format, as proportion correct choice (0.5 chance; 1.0 perfect discrimination). For 2D and S3D conditions, the mean proportion correct (squares), was plotted with ± 1 standard deviation as well as the maximum and minimum performance (downwards and upwards oriented triangles respectively). We applied a modified version of the ISO/IEC 29170-2 criteria where an image was considered to be visually lossless when the mean and standard deviation fall below 75%. For example, image F in Figure 2 is borderline when viewed in 2D and is visually lossless when viewed in S3D.

Data were analyzed statistically using a generalized linear mixed model (GLMM). The analyses revealed that, for the majority of the images, there was no difference in artefact detection rates for 2D and S3D viewing. In both of the experiments, where there were differences, artefacts were *less* detectable in the S3D conditions. These are images B. *RockMountain* (P = 0.0004) and F. *Aryaa* (P = 0.0001) in Experiment 1 see Figure 2, and Images C. *FluidB* (P = 0.0127), E. *FluidD* (P = 0.0050), and F. *FluidE* (P = 0.0484) in Experiment 2, see Figure 3.

Overall, our results show that for these images, which include examples with substantial differences in content in the two eye's views, artefacts are either equally visible in 2D and S3D or are less visible in S3D. This pattern of results is seen irrespective of whether performance for a given compressed image is visually lossless. For instance, in Figure 3 F the image is visually lossless in both 2D and S3D conditions, but observers are closer to chance in the latter. In Figure 2 B both conditions are lossy but again, detection rate is significantly lower in the S3D condition

4. Impact:

We have adapted the ISO/IEC Flicker paradigm to evaluate the visibility of compression artefacts using natural and computer-generated stereoscopic imagery with an aggressive level of compression. These experiments highlight the importance of image selection for stereoscopic testing. Our results show that even under high levels of compression, the visibility of compression artefacts is reduced for S3D content compared to corresponding 2D images.

5. Acknowledgements

The authors wish to acknowledge funding in support of this work received from the Ontario Centres of Excellence VIP II program and from Canada First Research to Innovation Funding to VISTA.

6. References

[1] N. Jacobson, V. Thirumalai, R. Joshi, and J. Goel, "A new display stream compression standard under development in VESA," Proceedings of SPIE 10396 103960U1-103960U12 (2017).

[2] R. S. Allison, L. M. Wilcox, W. Wang, D. M. Hoffman, Y. Hou, J. Goel, L. Deas, D. Stolitkza, "Large Scale Subjective Evaluation of Display Stream Compression" SID Symposium Digest of Technical Papers, 75(2) 1101-1104 (2017).

[3] A. Sudhama, M.D. Cutone, Y. Hou, J. Goel, D. Stolitzka, N. Jacobson, R.S. Allison, L.M. Wilcox. Visually Lossless Compression of High Dynamic Range Images: A Large-Scale Evaluation. SID Symposium Digest of Technical Papers, 49(1), 1151-115 (2018).

[4] D. Kane, P. Guan, & M. S. Banks. The limits of human stereopsis in space and time. Journal of Neuroscience, 34(4), 1397-1408 (2014).

[5] "Information technology — Advanced image coding and evaluation — Part 2: Evaluation procedure for nearly lossless coding," International Organization of Standards, Geneva, Switzerland, ISO/IEC 29170-2:2015, (2015).

[6] J.M. Harris & L.M. Wilcox. The role of monocularly visible regions in depth and surface perception. Vision Research, 49(22), 2666-2685 (2009).

[7] D.M. Hoffman & D. Stolitzka. A new standard method of subjective assessment of barely visible image artifacts and a new public database. Journal of the Society for Information Display, 22(12), 631-64 (2015).