Large Scale Subjective Evaluation of Display Stream Compression

Robert S. Allison[1], Laurie M. Wilcox[1],
Wei Wang[2], David M. Hoffman[2], Yuqian Hou[1],
James Goel[3], Lesley Deas[1], Dale Stolitzka[2]

[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]Samsung Electronics Co., Ltd., 3655 North First Street, San Jose, CA 95134, USA;
+1 (408) 544 4499; d.stolitzka@samsung.com

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

Abstract
VESA Display Stream Compression (DSC) is a light-weight codec designed for visually lossless compression over display links. Such high-performance algorithms must be evaluated subjectively to assess whether the codec meets visually lossless criteria. Here we present the first large-scale evaluation of DSC1.2 according to ISO/IEC 29170-2.

Author Keywords
display stream compression; visually lossless; subjective quality assessment; VESA; ISO/IEC

1. Objective and Background

VESA Display Stream Compression (DSC) is a lightweight video codec designed for low-impairment, low-latency compression and decompression over display links between computational devices and displays, or internal interfaces [1]. The codec is designed for a variety of display applications but in all cases the goal is to provide modest bandwidth reduction while maintaining sufficient fidelity so that a viewer cannot distinguish the compressed images or image sequences from the uncompressed source (visually lossless) [2].

The bandwidth of new image formats is outpacing the capacity of existing display links. Increasing the bandwidth by 2-4X using additional data lanes is undesirable due to complexity, pin count and system power. Real time compression algorithms could be a feasible alternative, however, compression is acutely lossy. Thus, there will typically be objective changes in the images which may be discernable by viewers. The distortions in such high-quality images can be quantified using various objective measures; however, no existing objective measure, or model of human vision, can accurately predict the visibility of artifacts in such low-impairment images [3]. Thus, subjective evaluation with inexperienced observers is required to validate that the compressed images are visually lossless.

This study set out to:

- Implement a large-scale subjective test with inexperienced observers, representative of the modern consumer market
- Verify DSC 1.2 visually lossless quality with a variety of types of difficult imagery to 8 bpp (3:1)
- Verify DSC 1.2 performance with new sub-sampled color modes, native 4:2:2 and 4:2:0, that are useful in consumer entertainment devices
- Implement a new motion panning test paradigm based on the ISO/IEC 29170-2 side-by-side protocol

2. *Methods*

2.1. Observers and Apparatus

Subjects were recruited from the York University community (N = 120) and were screened for color vision and acuity as specified in the standard [2]. No participants were excluded based on the pre-screening. Images were presented using custom Matlab scripts on an HP Dreamcolor Z24x monitor (1920x1200 @ 60 Hz, 30 pixels per degree at 45 cm viewing distance).

2.2. Stimuli

Twenty-five images (see thumbnails in Figure 1) were selected that included computer graphics, IT screens, people and animals, and challenge imagery [2]. DSC compression was applied to the full frame imagery and a 200 x 300 pixel region was cropped from both the original (reference) and compressed image. The location of the crop was selected to focus the observer's attention on a region of interest in the image.

2.3. Procedure

In the ISO/IEC 29170-2 protocols the viewer is shown two image sequences side-by-side, one of which contains compressed target images while the original is uncompressed. The standard recommends that the threshold for visually lossless performance be 75% correct for discriminating which of two images/sequences corresponds to the compressed image (performance midway between chance and perfect discrimination) although other criteria can be adopted [3].

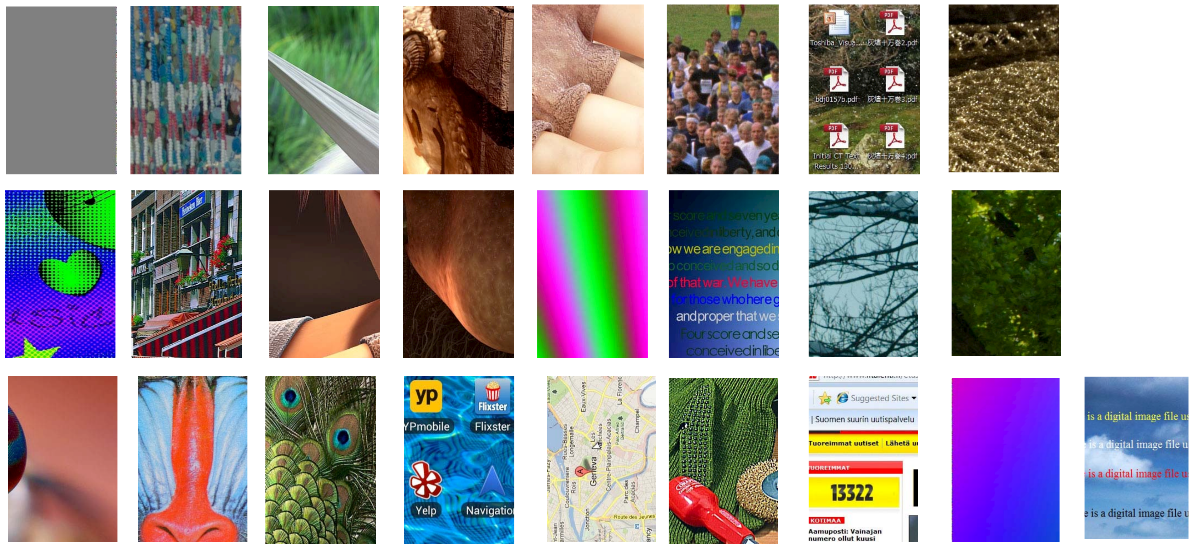

Figure 1. Set of 25 images used to assess DSC1.2. All images were cropped to 200 x 300 pixels following compression and displayed as described in the Methods.

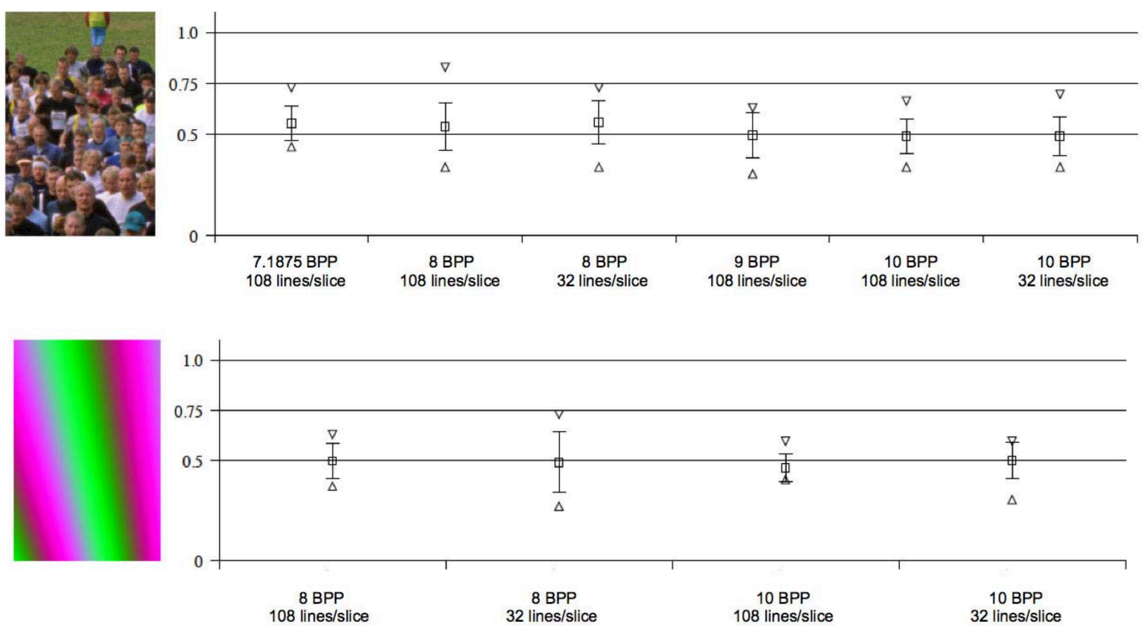

Figure 2. Sample proportion corrected identified for two images (shown to left, *Crowd* and *Starry Noise*) under a variety of evaluation conditions. Original content was 24-bit RGB format compressed to various bits per pixel (BPP). Square symbols represent the proportion correct averaged across observers. The error bars represent ± 1 standard deviation, and downward and upward triangles indicate the best and worst observed performance respectively.

We used two paradigms. In the ‘flicker’ protocol (Annex B from the ISO/IEC 29170-2 standard), two images (one the compressed ‘target’ and one the original) are presented side-by-side simultaneously, and each alternated with the original image at a rate of 5Hz. At this alternation rate the original-original pairing appears static. Given the flicker sensitivity of the human visual system [4], compression artefacts should be salient in the target-original alternation. On each trial the pair of sequences are shown for a maximum of 4s and the viewer is asked to indicate which of the pair contains flicker.

In the ‘panning’ protocol, (a proposed amendment to the standard) the compressed and original image sequences were again presented side-by-side. These side-by-side images both shifted 1 pixel in both the x and y dimensions every frame so the images diagonally panned back and forth across the displayed 200 x 300 pixel window (direction reversed every 1.5 s). One of the image sequences was compressed while the other was uncompressed. The observer's task was to report which sequence looked 'worse'. Observers viewed the sequences for a maximum of 10 s.

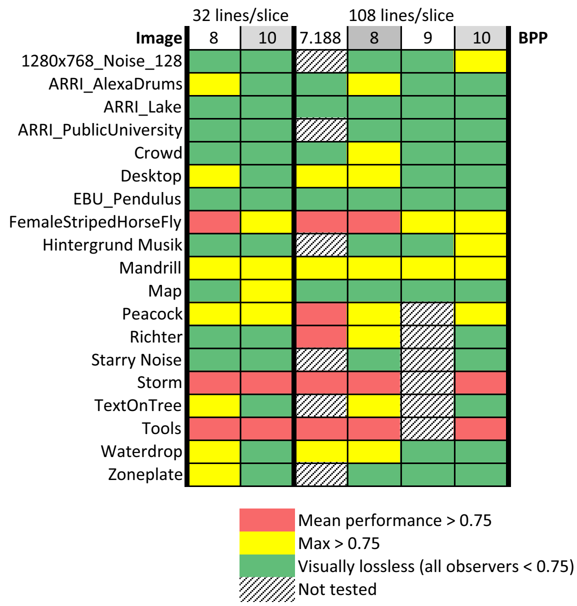

Figure 3. Flicker detection heat map for 24-bit RGB content compressed to various bit rates with slice heights of 32 and 108. Red indicates conditions where the *mean* detection rate exceeded 0.75, yellow where at least one observer had a detection rate of 0.75 or higher, and green where all detection rates were less than 0.75. Hatched spaces indicate conditions that were not tested.

2.4. Conditions

We compared flicker detection performance under compression at a comprehensive of a set of bit rates and for a variety of chroma subsampling formats (YUV444, 422 and 420). We also investigated the influence of DSC slice size. In the DSC algorithm slices are independently decodable rectangular regions that can be processed independently. DSC 1.2 is reportedly less prone to horizontal slice boundary artifacts than previous versions of the codec, therefore, slice width in these trials was always at least 2 slices per line to make it more likely that horizontal slice boundaries would be visible [1]. The slice height varied by either 108 lines or 32 lines per slice to test coding quality with respect to slice height due to uneven bit budget allocation in the first line. We might expect that artefacts would be slightly more visible at 32 compared to 108 lines per slice. However, Figure 5 shows that detection rate for the 108 line per slice conditions is generally similar to the 32 line per slice detection rate.

Catch trials were included with obvious flicker to monitor compliance with the task. Observers who could not detect more than 95% of catch trials were excluded. Panning sequence catch trials were also included but not used to exclude observers. All participants in the panning trials also completed static flicker testing and exclusion was based on static flicker catch trials.

3. Results

Descriptive statistics were computed from data for all sessions and blocks of trials and presented graphically according to the ISO/IEC 29170-2 protocol as proportion correct choice (0.5 guessing; 1.0 perfect discrimination). As illustrated by the sample data shown in Figure 2, for each condition, the mean proportion correct was plotted across observers with ± 1 standard deviation and symbols indicating the best and worst performing observers (downwards and upwards oriented triangles respectively).

Figure 3 shows the pattern of detection across the various bit rates for both 32 and 108 lin\es per slice for original content in 24-bit RGB format. As expected, the compressed images are more likely to be visually lossless (green) at higher bits per pixel (i.e., with less compression). The yellow highlighting indicates conditions where the most sensitive observer could detect the compressed image on more than 75% of the trials.

By the strict criterion used in ISO/IEC 29170-2 the conditions highlighted in yellow do not correspond to visually lossless behavior. Note, however, that use of the maximum individual response to determine visually lossless makes the classification sensitive to outliers. For example, in Figure 2 the 8 bpp results for 32 and 108 lines per slice are very similar, except that the maximum for the 108 case is greater than 0.75 and thus it is shaded yellow in Figure 3. This is despite the fact that all the other observers for this case would meet the visually lossless criterion.

The pattern of detection in the panning conditions is shown to the right in Figure 4. In general, most of the panning conditions were visually lossless. The decreased sensitivity to artefacts when the images are in motion is highlighted by the performance in the control conditions. The control images were intended to be readily detected, so the compression artefacts were obvious in the static images. However, our results show that for many images where detection was near perfect for the flicker control trials, they were degraded for the panning version.

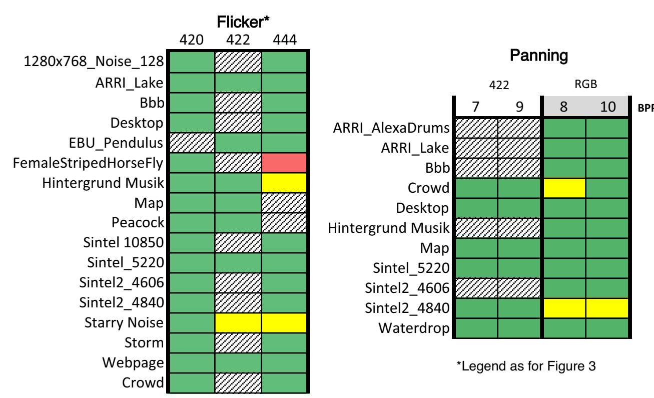

Figure 4. The heat map to the left shows flicker detection rates for YCbCr 444, 422 and 420 content compressed to 8 bits per pixel. The panning detection heat map to the right is for 24-bit RGB and 422 content. Cells are colour coded as in Figure 3.

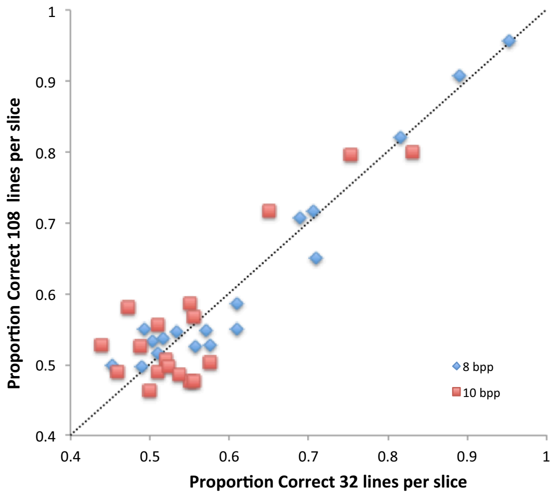

Figure 5. Each data point represents average detection performance across observers for a given image compressed with a slice height of 108 lines per slice plotted against performance on the same image with a slice height of 32 lines per slice.

4. Discussion

The current study is the first large scale evaluation of DSC 1.2 and follows evaluations of earlier versions of the protocol [3]. The codec was visually lossless on most images that represent typical and difficult image content down to 8 bpp or better for 24-bit 4:4:4 content. As expected, performance was generally better with equivalent 4:2:2 or 4:2:0 content. Performance on some images was not visually lossless, however, those images were challenging images with high entropy such as the *FemaleStripedHorsefly* (3rd row, 1st column Figure 1), *Storm* (2nd row, 6th column Figure 1), and *Tools* (3rd row, 6th column Figure 1) images. There were no obvious differences in sensitivity to flicker between images encoded with slice sizes of 32 line per slice and 108 line per slice.

In addition, we implemented a modified version of the flicker protocol using diagonally moving 'panning' versions of a subset of the images. The extension of the forced choice paradigm to include image motion is potentially important in that it is highly relevant to DSC use cases (e.g. mobile devices). The implementation of a paradigm containing the controlled panning motion is a viable amendment for the standard according to this study. Further, the perceptual impact of motion on perceived content quality is not straightforward. As pointed out by Choi, Cormack and Bovik [5], while setting content in motion may introduce new distortions (such as judder and blur), the spatio-temporal properties of the human visual system may help to make viewers less sensitive to flicker. Choi et al [5] go on to show reduced visibility of flicker in video sequences containing objects moving at relatively high velocities. They conclude that this effect is related to the phenomenon reported as 'motion silencing' by Suchow and Alvarez [6] and therefore its impact will likely be determined by the speed. To aid comparison with the flicker detection paradigm, in our panning trials, the stimulus layout was the same as that used described above for flicker, but within each stimulus window the image was shifted by 1 pixel in along the x and y axes; one of the image sequences was compressed, the other uncompressed. In this case the forced-choice task was to choose the image that looked 'worse'.

5. Impact

In this paper we evaluated the performance of DSC 1.2 under different levels of compression, different chroma subsampling formats, for a range of imagery, using ISO/IEC 29170-2 evaluation protocols. This was the first large scale evaluation of DSC 1.2 and the first to apply both the flicker and panning protocols. In most cases, visually lossless performance was achieved with target levels of compression for both RGB and YUV subsampled sources. This suggests the codec performs well with a wide range of content even under the stringent flicker paradigm. As suggested by psychophysical studies, the results of the panning paradigm implemented here show that in most cases when content is moving, observers are even less sensitive to compression artefacts.

6. Acknowledgements

This work was supported by Qualcomm, Samsung Display and VESA.

7. References

[1] F. G. Walls and A. S. MacInnis, "VESA Display Stream Compression for Television and Cinema Applications" IEEE Journal on Emerging and Selected Topics in Circuits and Systems **6**(**4**) 460–470 (2016).

[2] "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).

[3] D. M. Hoffman and 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**) 631–643 (2014).

[4] C. Landis, "Determinants of the Critical Flicker-Fusion Threshold" Physiological Reviews **34**(**2**) 259–286 (1954).

[5] L. K. Choi, L. K. Cormack, and A. C. Bovik, "Motion silencing of flicker distortions on naturalistic videos" Signal Processing: Image Communication **39**, **Part B** 328–341 (2015).

[6] J. W. Suchow and G. A. Alvarez, "Motion Silences Awareness of Visual Change" Current Biology **21**(**2**) 140–143 (2011).