# SUBJECTIVE ASSESSMENT OF IMAGE COMPRESSION ARTEFACTS ON STEREOSCOPIC DISPLAY

SANJIDA SHARMIN MOHONA

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

GRADUATE PROGRAMME IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
YORK UNIVERSITY
TORONTO, ONTARIO

SEPTEMBER 2018

# Abstract

Image and video quality are important to depict any pictorial information vividly and correctly. With the advancement of technology, we can produce high-quality images and can display those in advanced high-resolution displays. But as high-quality images continue to increase in size, transmitting these exceeds the limited bandwidth of display links. To cope, we need to compress the images but desire that the user cannot perceive any difference between the compressed and uncompressed images. In my thesis, psychophysical experiments with a flicker paradigm were undertaken to do a subjective assessment of the visibility of compression artefacts of two sets of images with two codecs viewed on a stereoscopic display. For one set of images the result shows that artefacts can be silenced in some stereo images relative to 2D while testing with the other set of images was inconclusive. This thesis documented evidence for silencing of artefacts in 3D displays. Other differences between stereoscopic and 2D presentation can be predicted but were not observed here (perhaps due to floor effects). Further large-scale subjective assessment with challenging images may help to get a concrete conclusion.

# Dedication

To my parents for their unconditional love

And

To my husband for his continual support

Thank you

## Acknowledgements

At first all the thanks go to the one and only Creator who creates the whole world and also myself and makes me able to finish my thesis as well as master's degree.

I want to give thank from the inner side of my heart to my supervisor Prof. Dr. Robert Allison. During all the way of my MASc study, he guided me to develop and present the ideas and solve any problem related to my courses and thesis. His consistent support gives me courage to go forward and successfully complete my master's degree. Getting a supervisor like him makes my MASc. experience more memorable and useful. I am thankful to him for his super patient attitude in answering all of my questions and removing my confusions. I appreciate our percept lab group, friends and colleagues for a wonderful two years research journey with them and also for their support and friendly behaviour. I am thankful to my committee members, York University, all the faculty and non-faculty members for their continuous support. Their presence, suggestion and encouragement help me to complete this work.

It was never possible for me to complete this thesis work without the support of my family. I am forever in-debt to my parents whose desire and support make me able to achieve this MASc. degree. I am also thankful to my parents-in-law for their continual encouragement and mental support.

Finally, I am grateful and thankful to my husband, Asad, who always supports me during the period of fear and frustration. His loving and caring attitude to me develop courage and confidence in myself. I am really very fortunate to have his support and company throughout my studies.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1
# Introduction

## 1.1 Motivation

Ultra high-resolution displays are becoming available and increasingly popular. With the increase of resolution, bandwidth demands for the video data interface increase significantly. Historically, the video source has transmitted uncompressed pixels to the display through the digital interface. But the bandwidth (BW) of such interfaces has not increased at the same rate as the growth in pixel count. So, maintaining a high data rate across the interface requires more power, more wires and more shielding to prevent electromagnetic interference (Walls and MacInnis 2016).These requirements increase device weight, hardware cost and complexity and are sometimes economically infeasible with current technology (Walls and MacInnis 2016).

Some existing algorithms including H.264 Intra-only, Motion JPEG 2000, Dirac/VC-2 can be used to reduce the BW but those are not designed to provide visually lossless quality with modest hardware complexity in real time, with the low latency expected of a display correction (Walls and MacInnis 2016) . To cope up with these problems, the industry has investigated lightweight, very low impairment compression techniques to reduce bandwidth requirements without impacting image quality. JPEG XS (Descampe et al. 2017), a ‘visually lossless’ codec, targets production applications, IP and ethernet application but it may not deal well

with specific types of content, e.g., subpixel rendered text (Walls and MacInnis 2016) . The Video Electronics Standards Association (VESA) introduced a standard for ‘visually lossless’, low cost and interoperable image/video compression designed to work in the display link that is now known as Display Stream Compression (DSC) (Walls and MacInnis 2016) and subsequently published another technique referred to as VESA Display Compression-M (VDC-M) (Video Association Electronic Standard 2018). In general, all practical image compression techniques are lossy, in the mathematical sense of being irreversible and thus losing information. However, this information loss may or may not be apparent to the viewer. Visually lossless implies that no visible changes are introduced in the compressed image compared to the uncompressed one. Visually lossless coded images or image sequences would be indistinguishable from the original under the same viewing condition and for the same spatial area (International Organization of Standards 2015) However, even in algorithms aiming to be visually lossless there may be some visible differences between the original and compressed images which are undesirable, distracting and may affect the subjective quality of image content. Image quality assessment is necessary to identify possible issues or artefacts produced by an image compression algorithm. DSC is claimed to be visually lossless up to a compression rate of 8 bpp (bit per pixel) (Walls and MacInnis 2016; Allison et al. 2017) .Objective measures of image difference such as S-CIELAB (Zhang and Wandell 1997) and Peak Signal-to-Noise

Ratio can be used for the assessment of a compression algorithm during development but the results from objective assessment do not truly anticipate human perception. Subjective assessment is essential to evaluate whether any algorithm, such as DSC or VDC-M, is visually lossless or not (Allison et al. 2017) as there is no perfect model of human vision to predict the visibility of compression artefacts (Hoffman and Stolitzka 2014) and humans vary in their sensitivity.

Objective and subjective measures for image assessment have been developed mostly in the context of 2D images and video. Using a stereoscopic 3D (S3D) display an observer can perceive depth or 3D structure from the information given by viewing with two eyes (binocular vision). In natural stereoscopic vision the depth information is signalled by the differences or disparities between the images in the two eyes. In an S3D display, such images are presented separately to the left and right eye and thus the user can perceive the depth by combining them in the brain. While objective and subjective measures for stereoscopic 3D image assessment have been developed (J. Yang et al. 2009; Wang et al. 2009; Campisi, Callet, and Marini 2007; Seuntiens, Meesters, and Ijsselsteijn 2006), to my knowledge no techniques for assessment of visually lossless stereoscopic images have been developed or assessed. In this thesis, I performed subjective assessment of DSC and VDC-M using stereo displays for both 2D and stereoscopic images using a variation of a standard technique for assessing whether an algorithm achieves visually lossless performance or not.

## 1.2 Contributions

In this thesis, I presented an experimental system based on the ISO 29170-2 flicker paradigm (International Organization of Standards 2015) in which 2D and stereoscopic images are displayed in a mirror stereoscopic display. The user's task was to select the compressed image using a gamepad when presented with a pair of images, one compressed and the other the uncompressed reference. This thesis contributes to the fields of subjective image assessment and image compression algorithms by:

i. Describing an experimental system that can be used to subjectively assess "visually lossless" image compression codecs. I also compared subjective with objective measures of images by calculating the Peak Signal to Noise Ratio (PSNR) (see Appendix).

ii. Comparing the effect of image compression codecs between 2D and stereoscopic images to find out if there is a difference in visibility of compression artefacts in 2D vs stereoscopic images. This has practical application for codec performance targets and algorithm development for 3D displays for gaming, movies, animation, 3D rendered images etc.

Based on theoretical considerations and experimental studies of human binocular vision, I can predict that the human perception of image artefacts may differ between 2D and stereo images under the same conditions. The objective of my

experiments is to assess the visibility of image compression artefacts in equivalent 2D and stereo images where equivalence is defined as:

i. Same display method and viewing conditions are used for both 2D and stereoscopic images.
ii. Same reference images are used for 2D and stereoscopic conditions.
iii. Same codec and same compression rate are used for both 2D and stereoscopic images.

## 1.3 Thesis Outline

The thesis consists of five chapters including this introduction. In the second chapter I review the literature on image compression, subjective assessment, stereoscopic vision and stereoscopic display, and binocular and monocular vision that are relevant to the assessment of compression artefacts.

The third chapter presents the general method and experimental setup. Image compression artefacts are assessed and compared in 2D versus stereoscopic images under the chosen test conditions. The experimental design and results along with statistical and theoretical analysis of these findings are discussed in the fourth chapter (1$^{st}$ experiment) and fifth chapter (2$^{nd}$ experiment). I summarize the conclusions and outline possible future work in this field in the final chapter.

# Chapter 2
# Literature Review

## 2.1 Review on Image Compression Artefacts

For transmission and storage of digital images, image compression has long been necessary. Image compression is a type of data compression, the purpose of which is reducing the size of an image or video file in a way so that the compressed file has an acceptable quality in-spite-of size reduction. For display link compression, the goal is to reduce the number of bits to be transmitted over the link. In other applications, the goal is saving more images in a given storage so that it may take less time to transfer compressed images through different devices and over internet, reducing the cost for storage (increase data capacity) and transmission-bandwidth (increase data rate).

In most images, unnecessary duplication of information occurs due to spatial and temporal correlation of neighbouring pixels. Due to this correlation images contain redundant information. Moreover, images contain information which is not perceptible to the human visual system (HVS) and this information can be considered as irrelevant. Image compression aims at reducing redundancy and irrelevancy (Dhawan 2011). Image compression can be classified as lossy compression or lossless compression. Lossless compression produces an image which is numerically identical to the original image. In lossy compression the

compressed image loses (bitwise lossy) information compared to the uncompressed one. This loss may produce visible artefacts. Lossy compression may be regarded as visually lossless if the difference (artefacts) between original and compressed image is imperceptible by the HVS. Some methods of lossless compression include entropy encoding, run length coding, chain codes, deflate and predictive coding; while color space reduction, chroma subsampling, transform coding and fractal compression are methods for lossy compression.

Lossy compression is more popular in some applications (e.g. streaming) as it can provide greater compression rates than lossless compression. But due to the irreversible nature, the concern with the lossy compression is the degradation of information which may result in poor quality i.e. visible compression artefacts. Lossy compression often includes quantization processes which reduce a range of values to a single quantum value. Block based coding used for quantization can introduce different types of artefacts including block boundary artefacts, ringing, posterizing and staircase noise. In block-based coding, each block is quantized independently which results different quantization coefficients between neighbouring blocks. Blocks containing an image edge can produce staircase noise where block bands appears like edge (Punchihewa and Bailey 2002). Ringing artefacts (Figure 2.1) can appear where sharp transition of signals occur. It appears as a ghost near the edges on a relatively uniform background due to a loss of high frequency components. In video, Mosquito noise arises from ringing in

successive frames, it appears as a shimmering blur of dots around the edge. Posterization (Figure 2.2) is an artefact which may occur due to color quantization when a continuous gradation of color tone cannot be sampled perfectly due to inadequate color depth. Posterization appears as discrete band of color in place of a continuous gradient of a color tone.

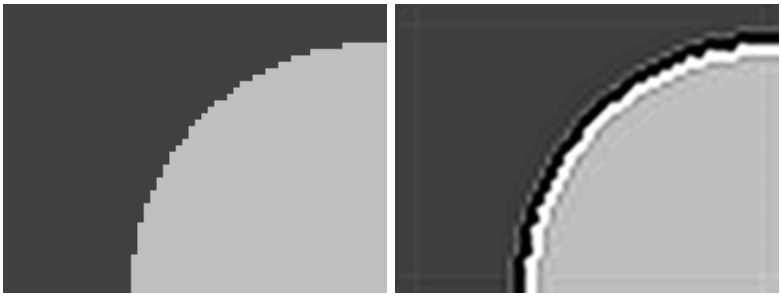


**Figure 2.1:** Image with (Right) and without (Left) Ringing artefact (Barth (CC0 1.0) 2010)

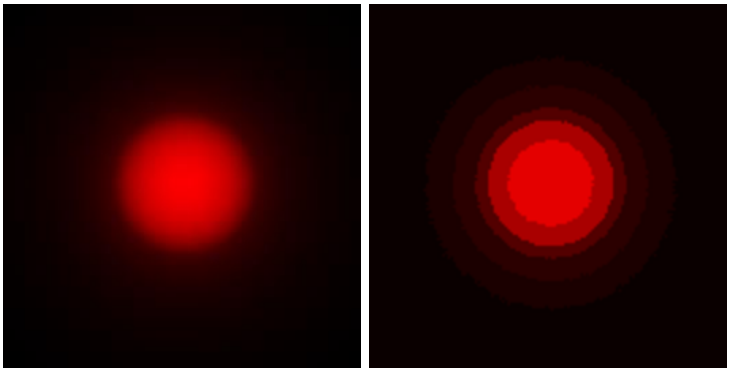


**Figure 2.2:** Image with (Right) and without (Left) Posterization artefact (Posterization Artifacts 2018)

As image compression is necessary and at the same time perception of artefacts is undesirable, so one of the main objectives of a compression codec is to provide visually lossless compression which can be evaluated by subjective assessment.

## 2.2 DSC and VDC-M

A range of different techniques have been developed to compress images and images sequences and given the range of constraints and application domains where compression is used there is no ideal codec and many codecs have been developed for niche applications. Historically one place that has always used uncompressed image sequences is the display link, the physical connection between the video source and the display. These display links, such as the video cable connecting your computer monitor, have always been intended to deliver the image with no distortion or loss of information. As discussed in the introduction there is now need for compression over this link, which needs to be performed with low latency, in real time and ideally maintaining the expectation of no degradation of the image (visually lossless). This compressed link is accomplished by an encoder-decoder pair or a codec. The purpose of the encoder is to take an input image stream and produce a compressed bitstream that is sent over the display link, while the decoder receives this bitstream as input and produces a reconstructed image sequence (Video Electronic Standard Association 2018)

To address these issues with display stream capacity, VESA developed a visually

lossless, low cost and interoperable image/video codec intended to work over the display link that is now known as Display Stream Compression (DSC 1.0) (Walls and MacInnis 2016). VESA released the current version of DSC, version 1.2, in 2016. Compared to other existing compression techniques, DSC provides visually lossless compression with low compression ratio, high data rate, constant bit rate, reduced complexity, low power and low cost (Walls and MacInnis 2016) . I used DSC in this study as it provides state of the art performance for display stream compression. Other popular techniques such as MPEG-2, H-2.64, JPEG-2000 and VC-2 require large storage which makes them too expensive and slow for display stream compression (Walls and MacInnis 2014) but they served as reference algorithms (Walls and MacInnis 2014) in developing DSC.

To be practical and generally applicable, objective and subjective assessment of a compression algorithm needs to be done for a wide range of photos and video, and for different kinds of displays to verify the visually lossless property. Besides a wide range of content from the target application domain, engineering ‘challenge’ images are often used that target suspected weakness based on the design and implementation of the codec or resource (memory, processing, …) limitations. Designing such images requires knowledge of the structure and operation of the compression technique.

The architecture of a DSC encoder and decoder is shown as a block diagram in Figures 2.3 and 2.4. In DSC, the pixels in each image are divided into an integer

number of non-overlapping, rectangular, slices, each of which is comprised of a set of groups where each group consists of 3 or 6 consecutive pixels. DSC uses delta size unit variable length coding (DSU-VLC) to code each group. The variable number of bits to code each group is converted to a constant bit rate by a rate buffer. An identical rate control algorithm is used in both the encoder and decoder and thus prevents buffer overflow and underflow. Subjective quality is maintained by the rate control algorithm as it adjusts the quantization parameter value. DSC uses a line buffer to store reconstructed pixel values from previous line. The decoder line buffer normally has the required storage to contain the full range reconstructed sample but if the decoder uses a smaller bit depth it may affect the picture quality. With RGB sources color space conversion is needed, DSC converts RGB source to YCoCg-R in the encoder and in the decoder YCoCg-R is converted back to RGB assuming spatial coherence of each component of the RGB signal. YCoCg-R is an alternative colour space that is better suited to compression and is a reversibly scaled version of the RGB space where Y represents the luma, Cg for chrominance green and Cr for chrominance orange. The colour space conversion is mathematically reversible for uncompressed images; however, for images where the RGB components are not coherent, e.g. images with chromatic aberration correction for a virtual reality headset, this compression of color space converted content may introduce artefacts.

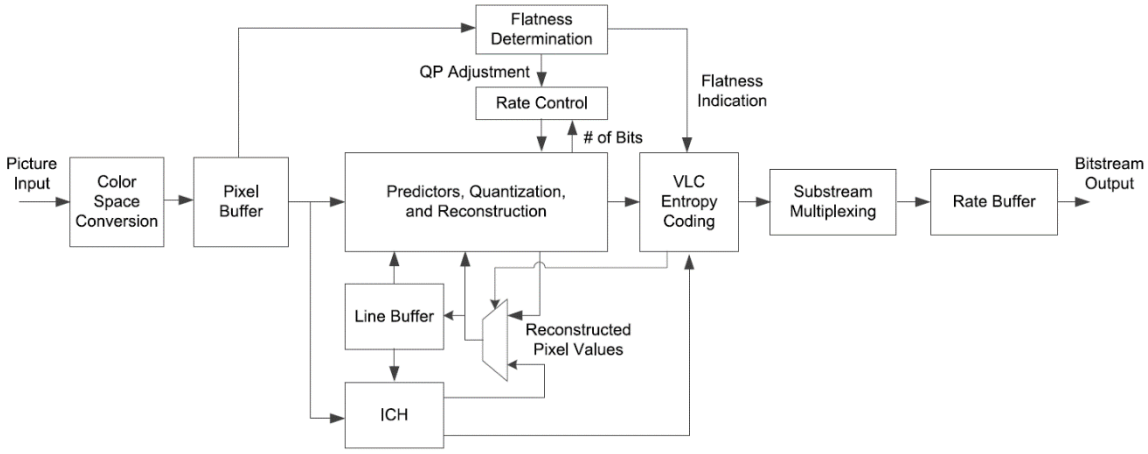



**Figure 2.3:** DSC encoder block diagram (Walls and MacInnis 2016)

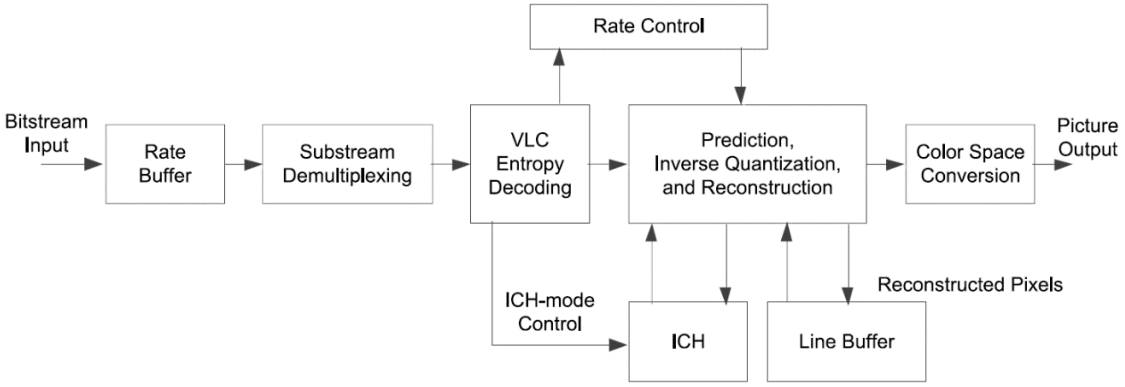



**Figure 2.4:** DSC decoder block (Walls and MacInnis 2016)

VDC-M was developed for mobile/VR/HDR applications to provide visually lossless compression of the display stream at higher compression rates than DSC. It was introduced to cope with the increase in pixel bandwidth for mobile and VR

applications, trading off a higher rate of compression by allowing increased computational complexity compared to DSC. Besides reducing the bandwidth required on the display link it also reduces the system cost by allowing a smaller frame buffer memory in the display driver IC (Figure 2.5).

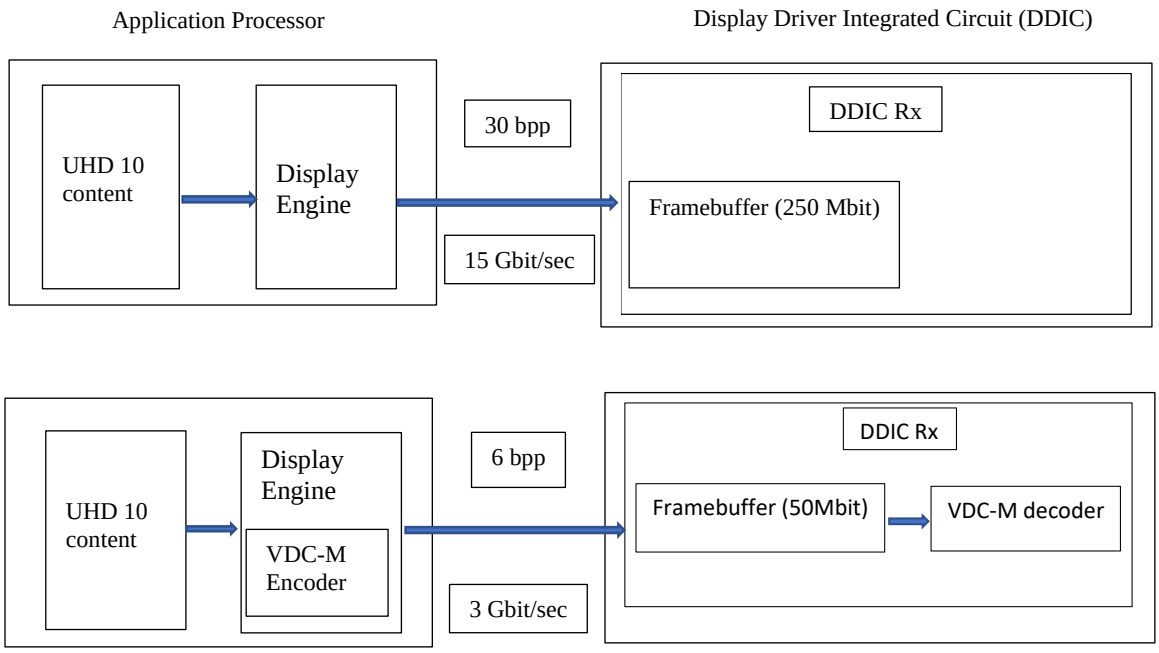



**Figure 2.5:** Example system configuration for a 30-bit, 3840 by 2160 pixel resolution display using VDC-M compression, adapted from (Jacobson et al. 2017)

Fig. 2.6 shows the functional block diagram of a VDC-M encoder. VDC-M is a real time, low bit rate compression technique which controls the underflow and overflow issues through the rate controller and buffer. When pixel data flow at a constant

rate to the encoder, the encoder produces a constant rate for the bitstream transmitted across the display link. As the budget and process technology of the display processor typically exceeds that of the display driver, the computational complexity was designed to be larger in the encoder than decoder. Through the bitstream syntax, the encoder sends the comparison results and mode decision to the decoder, allowing the decoder to be implemented with less complexity than encoder (Video Electronic Standard Association 2018).

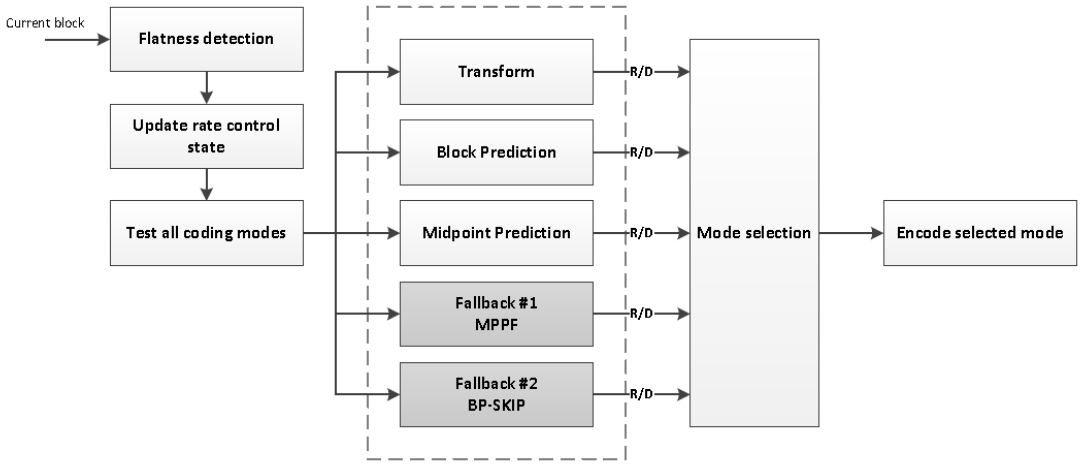



**Figure 2.6:** VDC-M Encoder (Jacobson et al. 2017)

I used VDC-M as well as DSC as it is also a state-of-the-art codec for display stream compression that allows higher compression ratios and involves different engineering trade-offs. As DSC or VDC-M are real time compression codecs where compression takes place in the display pipeline, they may result temporal and spatial image artefacts. Temporal instability or flicker artefact may appear in

this kind of real time compression as, with each screen refresh, a different image is compressed and decompressed. During subjective evaluation it should be remembered that this kind of flicker artefact may arise without any other artefacts visible in the individual images (Hoffman and Stolitzka 2014).

## 2.3 Subjective assessment of Image Compression

One major target in the image compression industry is to develop compression techniques which provide visually lossless distortion in maximally compressed images. Due to compression, artefacts may occur in the compressed image such as flickering, aliasing, color quantization (degradation of color quality), color banding, block boundary artefacts etc. Objective metrics can help detect some of these in the development of a codec but are not suitable to assess barely visible image artefacts in complicated images (Hoffman and Stolitzka 2014). For a visually lossless compression, the artefacts should not be discernible, and this must be determined or at least confirmed with subjective testing. The results of subjective image quality assessment, especially at visibility threshold, may be affected by various parameters such as measurement scale, industrial versus academic assessment setting, psychometric function assumed, and the subject's task. In many applications the measured threshold could be lower than that observed in everyday usage.

Zhang et al (Brunnström et al. 2017) identified three parameters that could affect

subjective testing: display, signal and viewing distance/angle. They did experiments to determine the effect of type of display on detection threshold at the same EOTFs (Electro-Optical Transfer Functions) and same viewing condition. For a digital display, the EOTF represents the mapping from the digital code words representing the content (“Electro”) to the visible light output on the display (“Optical”) (International Telecommunications Union 2018). This function has traditionally been described by the so-called gamma curve (International Telecommunications Union 2011) and is explicitly specified in newer high-dynamic range standards (International Telecommunications Union 2018; Society of Motion Picture and Television Engineers 2014). The EOTF affects the contrast and visibility of the artifacts and thus must be controlled for a valid comparison between displays. The authors measured contrast detection threshold of HEVC (High Efficiency Video Coding) distorted 8-bit images for two stimuli by three trained observers using a forced choice procedure. They tested observers on three different displays: a mobile display, a desktop, and a laboratory display. The displays were adjusted to have similar EOTFs and viewing distance was controlled to match visual angle. They found that contrast detection thresholds were similar for HEVC distorted 8-bit images in different displays when the EOTFs of the displays were similar.

Hoffman et al (Hoffman and Stolitzka 2014) introduced a new subjective method for assessing whether a compressed image was visually lossless subsequently

adopted as ISO/IEC international standard 29170-2 (International Organization of Standards 2015). The procedure was based on two-alternative forced choice detection where the uncompressed image and compressed image are both presented, and the test subject has to identify which is the compressed image. In such a scenario, if the compression is perfectly visually lossless the subject needs to guess, which results in correct responses on 50% of the trials on average. If the compression artefacts are obvious then they will always be seen (100% correct). When the artefacts are neither obvious nor imperceptible they will be detected with a probability that depends on their visibility. So, a statistical criterion or a difference threshold is needed to define the visually lossless property. This criterion is normally referred to as a 'just noticeable difference'(JND). JND is defined as the smallest difference between two stimuli that a user can detect some proportion of time. Hoffman et al (Hoffman and Stolitzka 2014) defined a threshold for visually lossless as one JND corresponding to a correct detection on 75% of trials under a set of reference viewing conditions, halfway between chance and perfect detection.

If the compressed and uncompressed images are displayed statically side by side, then it might be difficult to detect the difference. To make the artefacts more visible, Hoffman et al used a flicker viewing method in which two image sequences were viewed side by side for 4s. One was the reference image presented statically and the other was an alternation between reference and test image at a rate of 7.5 Hz.

When the compressed image is interleaved with the uncompressed one, differences between reference and compressed images flash (scintillate) and for this reason viewers can detect even subtle distortion which could be unnoticeable in a static side by side comparison (Hoffman and Stolitzka 2014). Additional temporal modulation occurs due to flickering that helps the viewer in searching the distorted region (Brunnström et al. 2017). Temporal frequency due to flickering can move the spatial frequency of the distortion to a more sensitive part of the spatiotemporal contrast sensitivity function compared to the static compressed image which is another reason why the artefacts become more visible with flickering (Brunnström et al. 2017).The visual system is sensitive to temporal change with a peak sensitivity in the 5 - 7.5 Hz range (Brunnström et al. 2017). Thus, the flicker viewing method makes the artefacts more visible. Instead of showing the whole image, a 200X300 pixel crop of each image containing the potential problem area was shown to ensure that observers focused on the area of interest. One criticism of the flicker paradigm is that it may be overly sensitive for testing still images (as still images do not flicker) as the flicker may produce visibility of artefacts not seen in still images. i.e., it may measure something which is irrelevant to still images (Brunnström et al. 2017). But for the compression codecs for display stream compression the expectation is very high that no artefacts are introduced during this compression stage, in this case the sensitivity

of the flicker paradigm is very conservative helping ensure the visually lossless property of this kind of codec (Brunnström et al. 2017).

Allison et al. (Allison et al. 2017) presented a large scale subjective evaluation of DSC 1.2 for 2D images at different levels of compression, different chroma subsampling and different slice sizes. They used the ISO/IEC 29170-2 standard with two protocols: Flicker and Panning. They used 25 images and tested 120 observers. In both protocols, a pair of image sequences were shown side by side: one of the images was the test image and the other was always uncompressed. The only difference between the two protocols was that in the flicker protocol the test image sequence consisted of the compressed image alternating with the uncompressed image while in the panning protocol the images were moved diagonally, and the test sequence was compressed while the other was not. The user's task was to identify which image looked worst or had flicker or artefacts. According to the standard if the correct discrimination of the compressed image in the pair of image was less than 75% for all observers then that compression was considered visually lossless (Hoffman and Stolitzka 2014). DSC 1.2 showed visually lossless performance for RGB and YUV subsampled sources at target levels of compression (down to 8bpp) but some challenging images were not visually lossless (Allison et al. 2017). Data from their psychophysical study shows that in most cases, for moving content in the panning protocol viewers were less sensitive to compression artefacts compared to the flicker protocol. This sensitivity

could be due to motion silencing (failure to detect any change due to motion). A subjective test was done in (Choi, Cormack, and Bovik 2015) to find out the effect of motion on detecting flicker distortion in naturalistic video which showed that the speed of object motion can affect the visibility of flicker distortion. According to (Choi, Cormack, and Bovik 2015) for objects with fast motion, subjects perceive less flicker distortion in spite of gazing on the moving object and in poor quality video motion silencing becomes more apparent.

Most of this previous empirical and theoretical work on image compression assessment has been performed using 2D images and video. Stereoscopic 3D image compression introduces additional considerations including the impact of artefacts on depth perception and the effects of symmetry of the compression artefacts in the two eyes. Chen et al (Chen, Bovik, and Cormack 2011) studied human detection of local distortion in stereoscopic 3D images. In the experiment they assessed perception of distortion by varying severity and type of distortion for low level image content in order to find out the effect of masking on stereo images. Four types of distortion were used: White noise, Blur, JPEG compression distortion, JPEG 2000 compression distortion. These distortions were inserted two ways into images. The first was as binocular distortion where the distortion was added in the same position in both images. The second was dichoptic distortion, where the distortion was inserted randomly in different positions in both images. The user's task was to detect the position of distortion. They found that for blur and

JP2k distortions in stereo-images, human perception was correlated with contrast and range variation; and regions with higher contrast or range energy were easily visible to user. On the other hand, contrast and range variation did not have significant effects for white noise and JPEG compression. These authors and others assessed binocular effects of noticeable image distortion and it is not clear if these results apply to near visually lossless compression where artefacts appear near their detection threshold. In this thesis, I compressed images using DSC or VDC-M codecs targeting visually lossless compression and determined whether threshold level of binocular or dichoptic artefacts are suppressed or enhanced by stereoscopic presentation. In analyzing these data, I seek to identify conditions where stereoscopic image compression may be more tolerant or conversely more susceptible to compression artefacts than equivalent 2D displays. To do so requires a basic understanding of human stereoscopic depth perception and S3D display techniques which are reviewed in the next section.

## 2.4 Stereoscopic Vision and Stereo Display

In stereoscopic display both monocular and binocular depth cues typically coexist (McIntire, Havig, and Geiselman 2014). The binocular cue of stereopsis arises from the fact that our eyes are separated laterally in the head so that each eye sees the world from a slightly different vantage point (binocular parallax). The differences between the images in the two eyes arise principally from depth

variation in the scene and are known as binocular disparities. When presented in an S3D display these left and right eye images are generally called a stereo pair. In a stereoscopic display, images are presented containing binocular disparity between the left and right images which gives the Human Visual System (HVS) information for relative depth perception. A S3D display must provide the appropriate image to each eye and there are different techniques available to separate and present the left and right views including mirrors, wavelength (anaglyph), polarization of light, alternating occlusion, and autostereoscopic displays. In a mirror stereoscope a pair of mirrors is used which are placed at ±45 degree angles to the face, one in front of each eye. The left side mirror reflects the left image and right side mirror reflects the right image, so each eye views the image designed for that eye and thus the user perceives depth by combining the two images in the brain. In the anaglyph technique, the left and right images are viewed through different color filters mounted in glasses, for example, one red and one cyan, and two images are printed overlaid on the same paper (or display) with the complementary colors. The anaglyph glasses can filter out the complementary color and can transmit the appropriate image to each eye. Polarizing techniques uses different polarizing filters to project two images on the screen. For stereoscopic vision the user needs to wear glasses with differently oriented polarizing filters (one matching each projected polarization) so that only similarly polarized light is passed through the filter and thus the two eyes have two different

images. Alternating occlusion methods use active shutter glasses synchronized with the display system. The glasses block one eye's view while presenting the image for another eye, e.g., when the right image is presented at that time the left eye is blocked. On the next video frame this occlusion is then swapped and the other eye's image displayed. This alternation is performed at a high enough rate that the user cannot perceive the repetition and can fuse the two images to perceive the 3D image. In an autostereoscopic display the user can perceive depth without the help of 3D glasses. For example, lenticular lens or parallax barrier technology can be used to direct the display of each half image in the stereopair while blocking the view of the other image.

Some basic phenomena that may arise and influence the visibility of compression artefacts with binocular viewing and stereoscopic display are listed below:

1. In stereoscopic images, when the size or position of object varies from one eye to other eye, it causes change in stereoscopic depth. Due to the compression artefact such differences may be produced leading to spurious binocular features. In stereopsis the compression difference in two eyes could appear as depth and the unseen pattern in 2D images can be disclosed. These depth effects can be generated due to the compression artefacts that make texture elements, appear, disappear, change shape or size.

2. When there are distortions in the shape, size or position of portions of one eye's image from the other eye, stereoscopic depth distortion can arise. This kind of depth distortion appears within an object or scene. Distorted depth occurs within the object and as a consequence shape distortion may occur. Generally global depth distortions are (Allison and Wilcox 2015) tolerable by users but DSC or VDC-M style compression may produce distortion and can result in shape artefacts which are generally more local.
3. When two images of similar pattern are shown to each eye, dichoptic masking arises. The detection of the test stimulus in the one eye is prevented by the 'mask' stimulus in another eye. When the images are similar, masking is most pronounced. Dichoptic masking should make the stereoscopic viewing more robust to DSC or VDC-M artefacts than 2D.
4. When a feature is introduced or removed in only one eye image, spurious monocular features arise. They are mostly seen at depth edges in an image. If the features are consistent with the viewing geometry, the depth change can be perceived. Spurious monocular feature can be introduced by DSC or VDC-M style compression as they may introduce or remove a feature or texture element.
5. Lustre/highlights on shiny surfaces can cause intensity variation in the two eyes. The shiny surface may appear as dull or a matte surface may appear as shiny when the intensity of a feature is affected by compression artefacts.

These artefacts may be more common in HDR (High Dynamic Range) because in HDR images the surface material properties can be represented with higher fidelity.

6. Accommodation-convergence conflict, excessive disparity (and depth) range in the images, poor stereoscopic image quality and other factors can introduce visual fatigue and discomfort. These factors are beyond the scope of the current thesis.
7. When sufficiently different images are presented to the two eyes that they cannot be 'fused' into a single image, binocular rivalry occurs. With binocular rivalry, at a given time only one image is 'dominant' or seen in any region of the display while the other eyes image is not seen or suppressed (binocular suppression). Any visually lossless or near visually lossless coding (or any modest to high quality coding) like DSC or VDC-M should not introduce such kind of extreme errors.

## 2.5 Binocular Vision vs Monocular Vision

When the same stimulus is viewed with two eyes rather than one, sensitivity to luminance, contrast, flicker, and motion increases. Since the stimulus appears 'stronger' the inputs are said to sum (Howard and Rogers 2012). This kind of phenomena is known as binocular summation. It could be affected by various factors. This happens when 2D images are viewed with both eyes instead of one

eye. This can be tested by displaying identical image in two monitors of a stereoscope. In terms of comparing S3D and 2D image compression, summation may be relevant in that the effects of noise on detectability of a pattern is larger when the noise is uncorrelated in the two eyes rather than correlated (Anderson and Movshon 1989). Thus, if independent compression noise arises in the two images it might be more apparent viewed stereoscopically. Probability summation is another type of binocular summation which has two subtypes: one with independent detectors and another with dependent detectors. When two different detectors with uncorrelated noise are presented with a weak stimulus rather than a single detector then probability of detecting at least one signal is higher. For perfectly correlated noise in two eyes, neural summation of two dichoptically presented stimuli has no advantage on Signal to Noise Ratio (SNR) as both the stimulus and noise signals are added. But in case of uncorrelated noise, neural convergence results in reduction in combined noise with better SNR. For monocular vision, neural convergence has no effect on probability summation as no independent decision process is taking place for monocular vision. According to (Campbell and Green 1965), convergence of neural signals from two eyes for two equal stimuli with equal and uncorrelated noise results in $\frac{2}{\sqrt{2}}$ binocular SNR which means that binocular sensitivity should be $\sqrt{2}$ times better than monocular sensitivity.

Campbell and Green (Campbell and Green 1965) used sinusoidal gratings of various spatial frequencies for the measurement of contrast sensitivity. They matched luminance and in monocular conditions one eye viewed a diffuse field of the same luminance. For the same spatial frequency in two eyes, the ratio of monocular to binocular contrast sensitivity was constant over the visible range of spatial frequency (Blake and Levinson 1977). To make a monocular grating equally visible as a binocular grating, 50% more contrast was needed (Legge 1984). This cannot be explained as doubling of luminance (summing the two eyes) as doubling the luminance of a monocular grating increased the contrast sensitivity by a ratio of only 1.17.

# Chapter 3
# General Method

To assess the effect of compression artefacts I ran a psychophysical experiment in a stereoscopic display to measure subjects' performance on detecting artefacts in 3D images and 2D images using a flicker protocol. The goal of the experiment was to determine (1) if there are significant differences in the visibility of compression artefacts in 3D compared to 2D images, (2) if there are significant differences in the visibility of compression artefacts when they are the same in the two eyes' images compared to when the compression differs in the two eyes and (3) to compare the compression artefacts for two different levels of compression for both 3D and 2D images. The compression algorithms used were DSC 1.2 (Experiment 1) and VDC-M 1.0.7 (Experiment 2) as representative state-of-the-art codecs designed to be visually lossless at modest compression levels.

## 3.1 Methods

In these experiments I made subjective assessments of the detectability of compression artefacts for stereoscopic images presented on a stereoscopic display. For the subjective assessment of image quality, test images were displayed to observers whose responses were used to determine the quality of codec, i.e. whether the artefacts were visible or not. In my experiment I used a version of the ISO/IEC 29170-2 (International Organization of Standards 2015)

procedure that I modified for stereoscopic displays. I used the “Flicker” protocol (Annex B from the ISO/IEC 29170-2 standard). In the 2D flicker method a pair of image sequences are displayed side by side where the pair consist of a test and reference sequence. The test sequence consists of the compressed image temporally interleaved (alternating) with the uncompressed one at a fixed frequency rate. I used 5 Hz as flicker rate, recommended in the ISO/IEC 29170-2 protocol, as this places potential artefacts near the peak of the spatio-temporal contrast sensitivity function (Brunnström et al. 2017). In the reference sequence the uncompressed image alternates with itself. When the images are interleaved the uncompressed-uncompressed sequence appears as static and the compressed-uncompressed sequence may show some artefacts due to the flicker sensitivity of Human Visual System. In the stereoscopic version of this task each image is an image pair consisting of a left and right half-image that combine to form a stereoscopic image when combined by the brain when viewed in the stereoscopic display. The compressed stereoscopic image consists of compressed half images (left or right or both) and the compression may be the same or differ in the two eyes leading to a variety of possible combinations of compressed stereoscopic images. In all cases though a compressed stereoscopic image alternating with the uncompressed stereoscopic image is the test sequence to be discriminated from a static uncompressed stereoscopic reference (Figure 3.1).

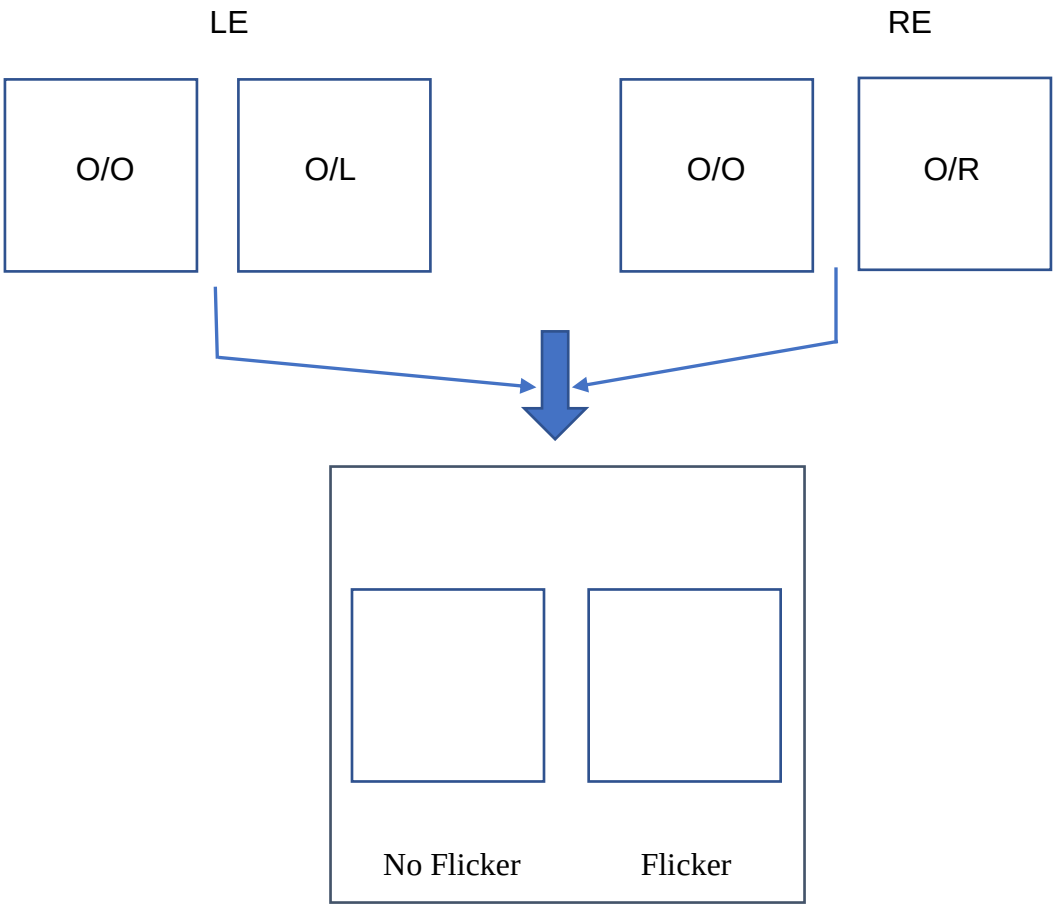



**Figure 3.1:** Display Protocol (O=Original image, L=Left test image, R=Right test image, LE=Left Eye display, RE=Right Eye display, O/O=Original image alternating with Original image, O/L=Original image alternating with left image.) The bottom shows the display as fused in the stereoscope.

**Figure 3.2:** Set up of experiment showing a subject detecting the flickered image by using a gamepad. The head is placed in front of the mirrors using a chin rest at a viewing distance of 45cm from each eye to its corresponding monitor.

In the experiment each subject had to perform a two-alternative-forced-choice-task on each trial. The task of the user was to detect the compressed image, that is, the flickered image. According to ISO/IEC 29170-2 protocol, for a particular image if a viewer could correctly detect the compressed image on more than 75% of trials then the algorithm was considered lossy for that image and viewer, which means that the artefacts were perceptible by the viewer. To monitor whether subjects could perform the task and were attending to the experiment catch trials were introduced using easily detectable control images. These control images were

highly distorted using the JPEG 2000 codec with compression quality 10. Observers data were included in the final analysis if they detected the control images on at least 95% of the trials. I ran two different experiments for two different set of images and two different codec (DSC 1.2 and VDC-M 1.0.7). Each experiment used a different set of 12 reference images, which were 2D images for Experiment 1 and 3D images for Experiment 2. For both experiments the task of user was to determine the flickered image from a pair of images of the same content displayed through the stereoscopic display.

### *3.1.1 Apparatus and System software*

A stereoscopic display was used to represent the stimuli. The stereoscopic display consisted of two HP Dreamcolor Z24x monitors arranged in a mirror stereoscope configuration, mirrors were placed at a 90 degree angle to each other, ±45 degrees angle to the user's face. The screen size was 51.5 cm X 32.5 cm, pixel resolution was 1920 X 1200 and the frame rate was 60Hz. The monitors were carefully matched for color and luminance (maximum 119 $Cd/m^2$ and minimum 0.14 $Cd/m^2$). The viewing distance was 45 cm (30 pixels per degree) from the chin rest placed in the middle of the two monitors. Participants held the gamepad (Microsoft SideWinder Plug & Play Game Pad) with two hands and provided their response by clicking one of the two shoulder buttons. The left shoulder button was used to select the left image and right shoulder button to select the right image.

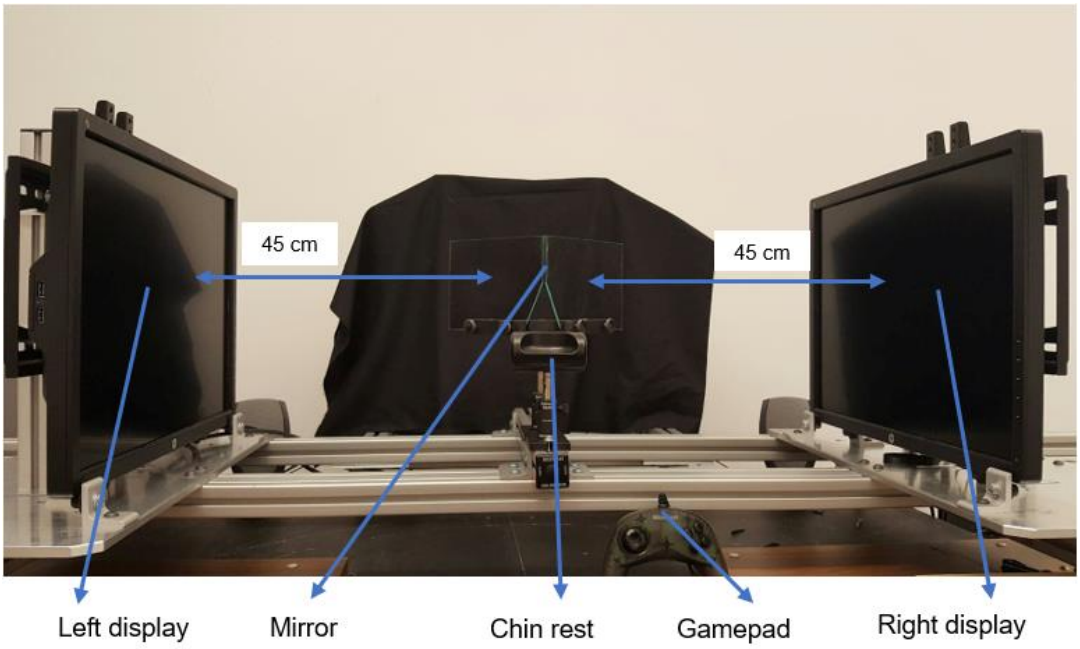



**Figure 3.3:** Test Station

The Psychotoolbox package (http:psychtoolbox.org/) running in MATLAB R2016a (https://www.mathworks.com/) was used to develop the script for the experiment code, to display the stimuli and to record users' responses. For statistical analysis, R version 3.4.2 (R core Team) was used.

### *3.1.2 Participant Screening*

To be qualified for participating in this experiment each user had to pass three types of vision test: one for color vision, one for stereo vision (observers must have the ability to discriminate disparity of 40 seconds of arc or better) and one for visual acuity (observers must have better than 20/20 corrected acuity). Participants also had to between the age of 18 to 40 years as per the standard.

To measure visual acuity a Snellen chart was used. In a Snellen chart the measure of normal vision is referred to as 20/20; which implies that the observers can identifying optotypes nominally sized for a 20' viewing distance at an actual distance of 20'. A participant with 'normal' visual acuity can identify an optotype which subtends 5 minutes of arc and has 1 minute of arc features and these are presented on the 20/20 line. The passing score to be included in the experiment was 20/20 with habitual visual correction.

To measure participant's depth perception ability, stereoscopic acuity (stereoacuity) was measured. The smallest detectable depth difference perceived by binocular vision determines the stereoacuity. For my experiment, the Randot stereo test (2015 Stereo Optical Company, Inc.) was used, which is a vectograph test. It can measure stereoacuity from 400 arc second down to 20 arc second. The passing score for the participants to be included in the experiment was 40 arc second. During this test the participant had to wear polarizing glasses, could not tilt their head to the side and the distance of test plate from eye was 16 inches. Adequate light was provided for the test but reflection was avoided from the surface of the test booklet.

To check for color deficiency the Ishihara test (Kanehara Trading Inc., 24 plate edition, 2005) was used which contained a series of pictures composed of colored spots to diagnose red-green color deficiency. During the test, the test plates were placed at 75 cm from eye so that the test plate was at right angle to the line of

vision. Six plates were used to test the participants and they needed to provide correct answers for all 6 plates with no more than 3 s delay to be qualified for the experiment.

# Chapter 4
# Experiment 1: 2D vs 3D using 2D images

In Experiment 1, we used 2D images and assessed the effects of (1) disparity of the images relative to the screen and (2) the matching or symmetry of the compression errors in the left and right images. This allowed us to use images that were previously used for extensive 2D testing of the DSC protocol and this previous dataset served as a baseline for comparison for our testing in an S3D display. More importantly, the use of 2D images with disparity allowed for assessment of the effects of displaying the images with a constant disparity with respect to the screen (a disparity pedestal) without the complications of variation in disparity within the image as would be present in true stereoscopic images. This allowed for assessment of the effects of disparity pedestal and to investigate whether there were images for which the performance of DSC 1.2 was different between disparity and no disparity conditions. The use of 2D images also allowed for assessment of the effects of compression artefacts when they were identical in the two eyes' images compared to when they differed, by using identical 2D images in the left and right eyes the image compression in the two half images could be matched perfectly. Alternatively, by offsetting the images before compression, different compression errors could be introduced in the left and right images. This experiment used 2D images presented as stereoscopic pairs to

address these research questions. Experiment 2 extended these measurements to consider true stereoscopic 3D content.

## 4.1 Methods

### *4.1.1 Subjects and Conditions*

Sixteen participants were recruited for Experiment 1, but after visual screening four participants were excluded (two for poor stereoscopic vision and two for poor visual acuity) and one participant was excluded for failing to detect flicker in control images. So, finally eleven subjects were tested and their data included for analysis. Among these, 8 participants were female, and 3 participants were male (ages of participants: 18 to 29). All the participants were from the York University community. Two of them were York University Graduate Students, one participant was an Undergraduate Research Assistant and the remaining 8 participants were signed up through the York University Undergraduate Research Participant Pool (URPP). For participating in the experiment 1 credit/hour was given to each of URPP participant and other users were paid at $20/hour.

The codec used in this experiment was DSC 1.2. Image pairs consisted of identical left and right half images that were presented (1) at the same location in both left and right displays so that they had zero disparity and appeared at the plane of the screen or (2) where shifted in opposite direction on the left and right displays to produce a disparity relative to the screen and appeared to lie behind the screen.

There was no disparity variation within the images themselves only an overall shift thus the images appeared as a 2D image lying either in the screen plane or offset from it (see Experiment 2 for true stereoscopic test images). The use of identical left and right half images allows for introducing identical image compression in the left and right eye views, which would not be possible in true stereoscopic image pairs (as the source images themselves differ). The compression introduced in left and right half image could either be the same or different. Thus, there were four conditions used for this experiment. These are as follows:

1. Disparity with same compression in the left eye and right eye image.

2. Disparity with different compression in the left eye and right eye image.

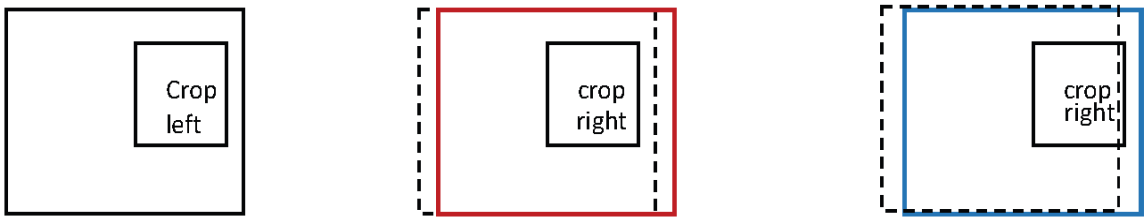
3(a). No disparity with same compression in the left eye and right eye image using only the left half image from condition 2 for both eyes.

3(b). No disparity with same compression in the left eye and right eye image using only the right half image from condition 2 for both eyes.

4. No disparity with different compression in the left eye and right eye images.

All conditions were dichoptic and the second condition corresponded to the typical compressed stereoscopic display case. In my experiment, I used a selected (i.e., cropped) portion of image instead of a full image. The disparity used for conditions 1 and 2 was 10 pixels. Therefore, to prepare 3D images I did 10-pixel horizontal shifting on the 2D images after and before image compression for same

compression and different compression, respectively. The image shift prior to compression caused the images to differ and produced small stochastic variations in the compressed images even at the same compression levels.




Original image → compressed Original image shifted → compressed

Crop region defined 1.Crop is not shifted 2.Crop is shifted

(have disparity) (no disparity)

**Figure 4.1:** Dichoptic Combinations

For condition 1, shifting was done after compression to ensure the compression was the same in both eyes' images. The full-size compressed image was cropped for the left eye image. Then, to introduce disparity into the stereo pair, the full compressed image was shifted but the crop position was not shifted to produce the right eye image. Thus, I obtained the compressed left eye and right eye images for condition 1. To produce the reference (i.e. uncompressed) image pair for condition 1, the original image was cropped to produce the left eye image and then to produce the right eye image, the original image was shifted and cropped in the same position as for the left eye image.

In case of condition 2, the original image was compressed and cropped to produce left eye image; for the right eye image, the original image was shifted before compression, the image was compressed and after compression the image cropped in the same position as used for left eye image. Thus, the displayed image pair has disparity and different compression in the two eyes. To prepare the reference image pair of condition 2, the original image was cropped to produce the left eye image and the right eye image was prepared by shifting original image and then cropping in the same position as for left eye image.

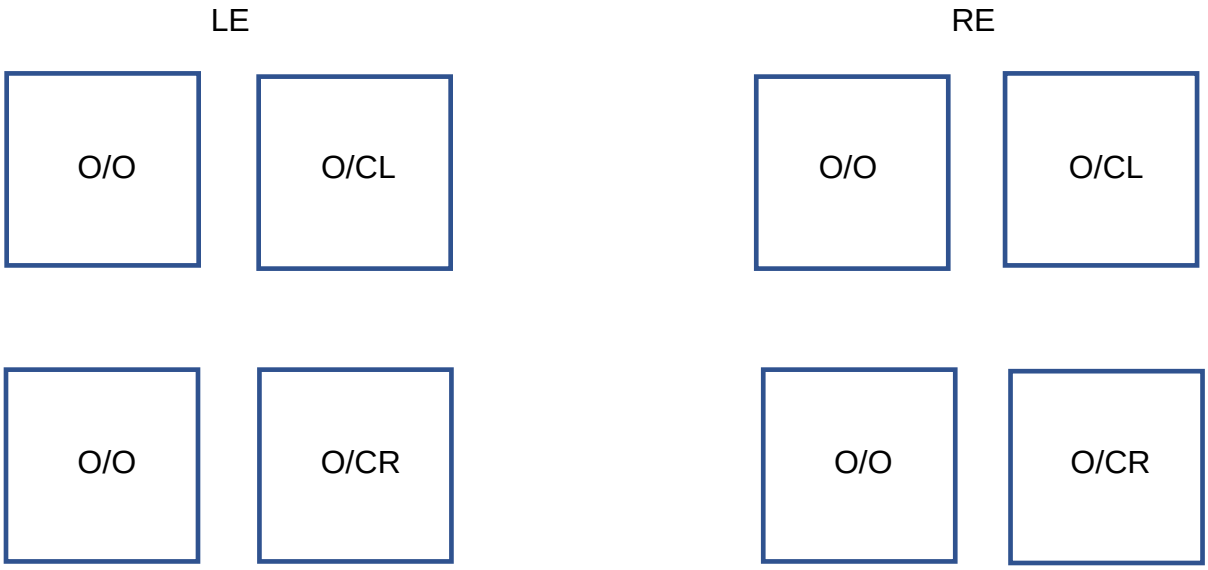



**Figure 4.2:** Binocular Combination (O=Original; CL, CR=Compressed, cropped Left and Right image; LE= Left Eye monitor; RE=Right Eye monitor). Top shows Condition 3a while bottom shows Condition 3b

Condition 3 was displayed as in Figure 4.2, either the left (compressed and cropped) or right image (compressed and cropped) was displayed to both eyes so

the displayed image had no disparity and the same compression to both eyes. For this condition the compressed left image was obtained by cropping the original image after compression and the reference left image was obtained by cropping the original image before compression. To get the reference right image, the original image was cropped after shifting 10 pixels horizontally. To get the compressed right image, the original image was shifted 10 pixels, then compressed and cropped. Both the compressed left and compressed right images were used in case there was a difference in artefact visibility in the two cases due to the image shifting.

In the fourth condition, the original image was compressed and cropped to give the left eye test image; then for the right eye test image the original was shifted and then compressed so images have different compression but the crop was shifted by the same amount as the image shift, so it created zero disparity. The left eye and right eye images for the reference pair were the same image, which was produced by cropping the original image before compression at the same location.

### *4.1.2 Stimuli and Experiment Procedure*

For experiment 1, I used a set of twelve 2D images from the VESA test data set that were particularly difficult to compress in previous VESA testing and hence nicknamed the ‘killerImage’ set. In the selection of a set of images, it is necessary to find out the artefact location and the likelihood for a given image to be lossy or

lossless after applying the codec. Computer analysis can be used to narrow down the search for the artefact location. In the process of selecting the images, VESA used a computer-aided analysis which helped to reduce the image candidates for the flicker paradigm experiment (Stolitzka 2017). This process started with an initial set of 10,000 still images and image frame sequences contributed to VESA. These 10,000 images were compressed and filtered for potential artefacts using some common techniques including PSNR, SSIM, SCIELAB image mapping and an image's history. From these filtered images (2nd set) a 3rd set of images were selected for the flicker paradigm test; *Barbara, Zelda, Peppers* and some other images from the Kodak 24-image test set were included in this set. Assessment of human experts was used during the selection of the 2nd set of images so that the set contained various types of image scenes including still life, web capture images, human portraits, screen backgrounds, landscapes and animals. The images which had high PSNR value (>60db) were considered as lossless (too easy for the codec) and excluded during selection process but some unique images were included in the 2nd set in spite of high PSNR if there were expectations of local image issues or they were representative of a class of images. In this process mostly lossless and unexciting images for users were excluded. Following the above process 200 images were selected using computer-aided analysis from the 1st set of 10,000 images. Then 50 images were selected from 200 images through image mapping and scene selection. Image mapping categorizes images and

scene selection makes sure that representative images are included. In the image mapping process, the authors labelled the images into a variety of categories (still life, web capture images, human portraits, …) and selected from these to include representative images from a variety of categories. From these 50 images, a set of 12 images (Hoffman and Stolitzka 2014), difficult for the codec to compress, were used in the experiment as shown in Figure 4.3. The flicker paradigm was applied on these full-sized images and the locations of artefacts were identified by human experts. The original 1920 x 1200 pixels images were cropped to 200 x 300 pixels to isolate these regions containing the artefacts. Original images were 24-bit full colour standard dynamic range RGB images (8 bits per pixel per colour channel (bpc), 24 bits per pixel (bpp)) that were compressed 3:1 by DSC 1.2 to produce 8 bpp images. For display, these images were decompressed to produce 24-bit RGB images containing any artefacts produced by the compression step.

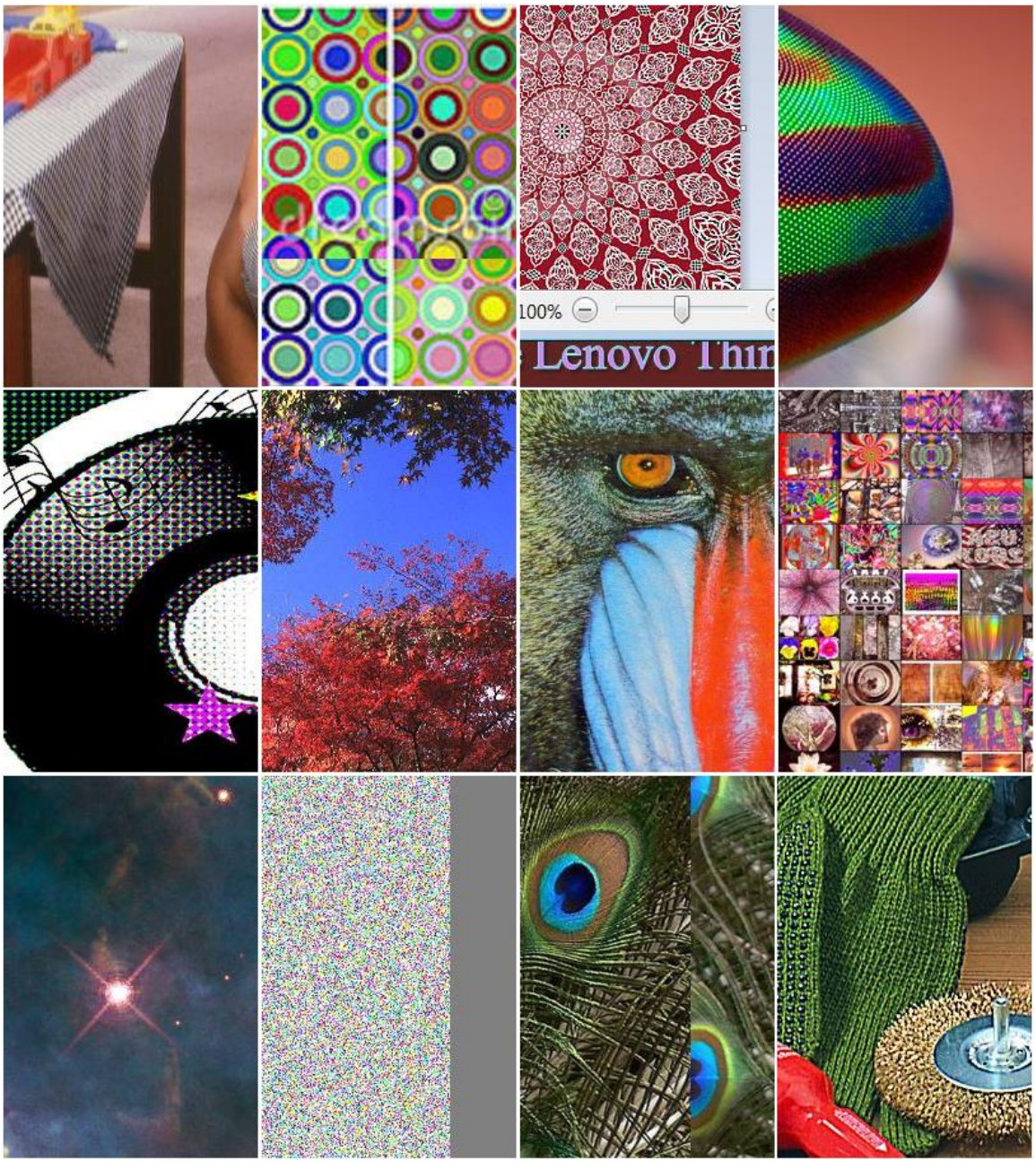



**Figure 4.3:** Twelve images for experiment 1

As described above, four main conditions were tested with the third condition divided into two sub-conditions: one that showed compression based on the left images and the other based on the right images. For each condition the same set of 12 images was used and 6 catch trial images were also added to each session. Performance on these catch trials indicates understanding and attention to the task. Each of 12 the images in each condition was presented for 20 times resulting in 240 trials per condition (1200 trials for 5 conditions) and each catch trial image was presented 20 times in each session resulting in 120 catch trials per session (total 240 catch trials in 2 sessions). The total number of trials per observer was 1440 including the catch trials and viewing time for each trial was 4 seconds. Participants indicated whether they saw flicker in the left or right in a pair of stereoscopic images by pressing one of two shoulder buttons on the gamepad. They could provide their response during the 4 s trial and a feedback sound was provided when they failed to select the target image. If they did not respond during the trial a blank screen appeared with a text message asking the subjects to provide their response.

The whole experiment was performed by each participant in two sessions over two separate days, 60 minutes in each day. Prior to the experiment, each participant provided written informed consent to participate in the experiment and was screened for visual acuity, stereo acuity and color acuity. This process took about

10-12 minutes. Following screening, the first day's session was run. To limit user fatigue, each session was divided into blocks so that each block time could be competed 5-6 minutes. The session was divided into blocks of trials for three of the conditions, every condition-image combination was tested twice in a random sequence to the user in each of the 8 blocks tested on the 1st day. On the second day the 3 conditions tested during the first day were again tested through another 2 blocks (to complete the 20 trials per image-condition) and then the images for other two conditions were tested in random sequence through 10 blocks with 2 trials per image per condition in each block.

## 4.2 Results

In this experiment, the primary hypothesis to be tested was that the sensitivity to compression artefacts differed between conditions with disparity and conditions without where the stimulus appeared at the screen plane (Hypothesis 1.1). The second main hypothesis was that sensitivity to compression artefacts would differ when the compression was symmetric (same) versus asymmetric (different) in the two eyes (Hypothesis 1.2). To assess these hypotheses, we need to compare relative performance between the different conditions when images are compressed using DSC 1.2.

### *4.2.1 Descriptive Results*

The average proportion correct score for each condition with DSC1.2 compression is shown in Fig 4.4,4.5 and 4.6. This figure shows the data from subjects who were able to detect the control images with a proportion correct greater than or equal to 0.95. For each condition, the mean proportion correct score for the subjects is plotted with ± 1 standard deviation (SD). The square symbol represents the mean, the error bar represents the standard deviation and triangles denote the range of scores. The best performing observer (for whom the proportion correct score is highest, i.e., who can detect the flicker most often for a given condition) is indicated by downwards triangle and worst performing observer (for whom the proportion correct score is lowest) is indicated by upwards triangle. According to ISO/IEC 29170-2, if any subject can detect the flicker in image for more than 75% time then it can be said that DSC1.2 is lossy in that condition. But this definition is based on the best performing observer which makes the classification sensitive to outliers. Thus an alternative criterion that is less sensitive to outliers was adopted: the compression was considered visually lossless if the mean detection rate plus 1 standard deviation was less than the 0.75 criterion (Allison et al. 2017). The main objective of this study was not to verify visually lossless performance but to measure relative performance between conditions. We cannot do this comparison directly in Fig 4.4, 4.5 and 4.6 because of the asymmetric nature of underlying

binomial distribution and because the repeated measures nature of the data is not considered.

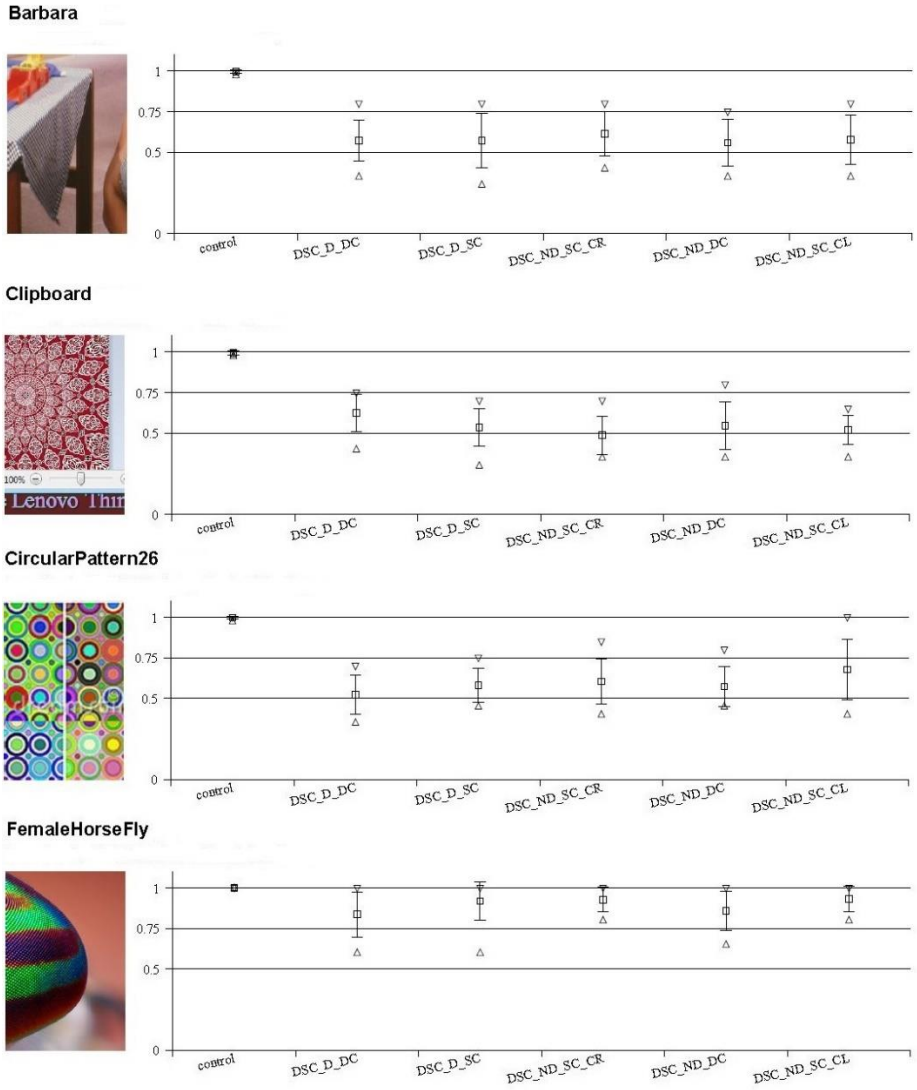



**Figure 4.4:** Sample proportion correct under different conditions with DSC1.2 Compression (part 1) (D = disparity, ND = no disparity, DC = different compression, SC = same compression). Error bars represents ±1 SD.

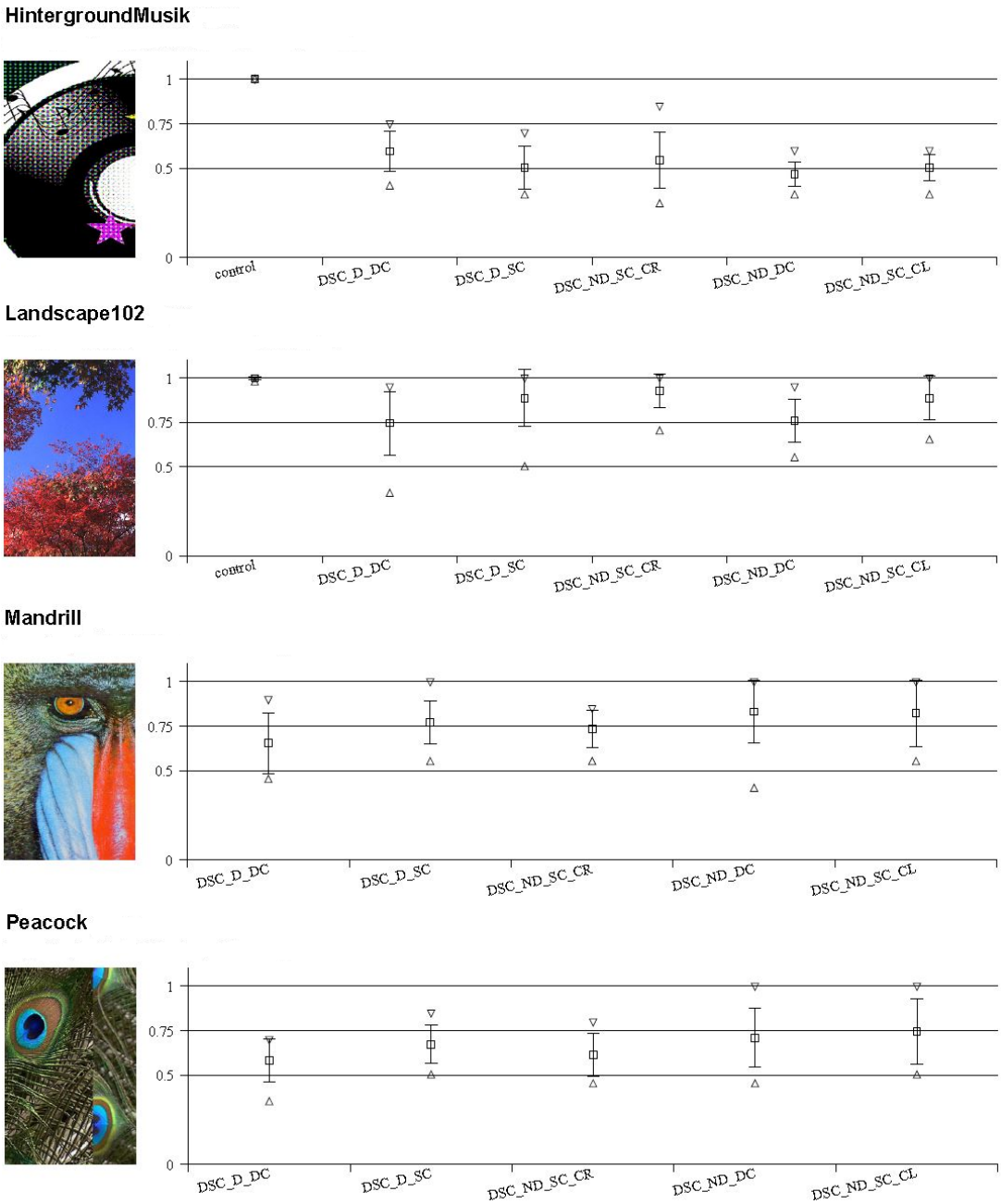



**Figure 4.5:** Sample proportion correct under different conditions with DSC1.2 Compression (part 2). Error bars represents ±1 SD.

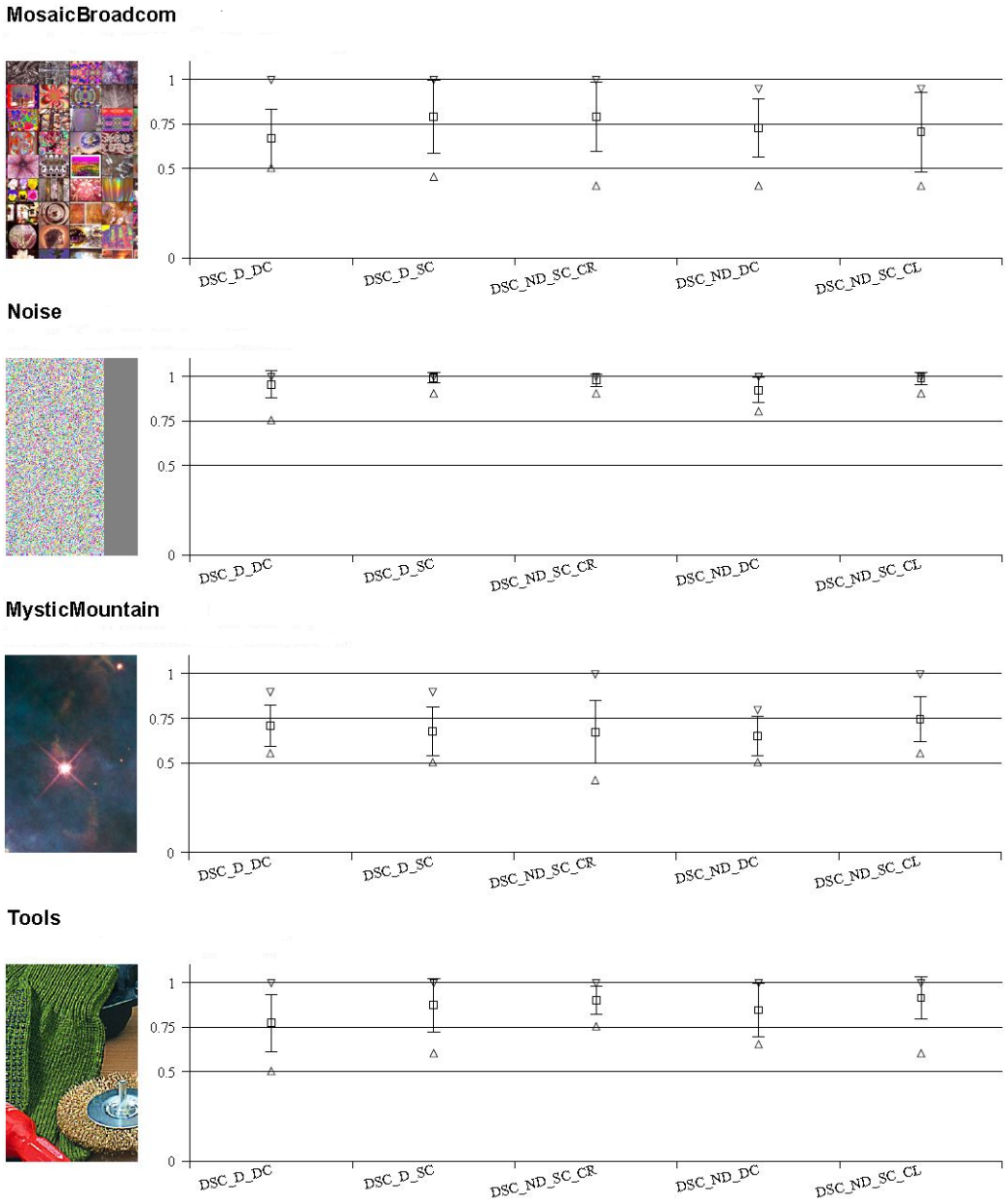



**Figure 4.6:** Sample proportion correct under different conditions with DSC1.2 Compression (part 3). Error bars represents ±1 SD.

### *4.2.2 GLMM analysis*

To test the two hypothesises, the flicker paradigm data was fitted using a Generalized Linear Mixed Model (GLMM) (Agresti 2007). GLMM is an extension of GLM (Generalized Linear Model) and GLM is a generalization of SP Model (LM). Generalized linear models allow for response variables that have distributions other than normal through specifying appropriate link functions between the linear predictor and the response variable. In this case the responses were binary choices and the binomial distribution, and a logit link function were used. Its linear predictor includes both fixed and random effect. LM allows for only normal distribution but GLM allows both normal and non-normal distribution and assumes that data are independent. GLMM allows for non-independent in addition to other features of GLM. Sometimes data are not independent e.g., when each participant provides more than one data point due to repeated measurement under different conditions. As the collected data of my experiment has non-normal distribution and non-independent nature due to repeated measurement, so I used GLMM for the data analysis.

GLMM analysis was done in the R statistical software environment (R Core Team 2017) and maximum likelihood estimation was used for fitting data using lme4's 'glmer'. Before fitting GLMM, control data was excluded because including control data can cause convergence failure as the response proportions were all 1.0 for control data.

My main hypotheses were that there would be significant effects of both depth and compression. Depth has two levels: disparity (3D or stereo, Conditions 1, and 2) and without disparity (2D, Conditions 3a,b and 4); whereas compression also has two levels: same compression (Conditions 2 and 4) and different compression (Conditions 1, 3(a), and 3(b)). To assess the effects of depth, compression and their possible interaction on artefact visibility I used the following formula for the model:

correct ~ depth * compression + (1 | subject) + (1 | ref) (1)

The model is expressed in the so-called Wilkinson notation (Wilkinson and Rogers 1973) for describing statistical models (this notation is also used for model specification in a variety of software packages including Matlab, S-plus, and R). In the above formula the measured response variable is on the left-hand side of the '~' operator and the right hand side describes the model in terms of the predictor variables or factors. Terms are added to the linear model with the '+' operator and '*' represents both main effects and interaction of the factors, that is depth*compression is equivalent to depth + compression + depth : compression where the ':' operator specifies the interaction between the two factors. For the above formula, the model treats depth and compression as fixed effects. I modeled both subject ID and image as random factors which is indicated by the grouping notation (1|subject) which indicates that a random intercept model is used for the within subject variable. The error distribution of the response variable was modeled

as binomial. The summary of GLMM fit (using ‘glmer’ in R) for the above formula is in Table 1.

**Table 1:** Effect of interaction between depth and compression for 1st set of images (Variables: 2 depth conditions (stereo, 2D), 2 compression conditions (same, different), 11 subjects, 12 reference images)

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 1.193 | 0.29 | 4.09 | 4.40e-05 *** |
| depthstereo | -0.227 | 0.06 | -4.08 | 4.43e-05 *** |
| compressionsame | 0.030 | 0.06 | 0.53 | 0.594 |
| depthstereo:compressionsame | 0.219 | 0.09 | 2.53 | 0.011 * |
| Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 | | | | |

From the above result we can see that, across all subjects and images, the interaction of depth and compression had a significant effect on the perception of compression artefacts. On average artefacts are less visible, when compression is the different in the two eyes than for the same compression (Figure 4.7) but this is particularly true for images with disparity which explains the interaction (Figure 4.8)

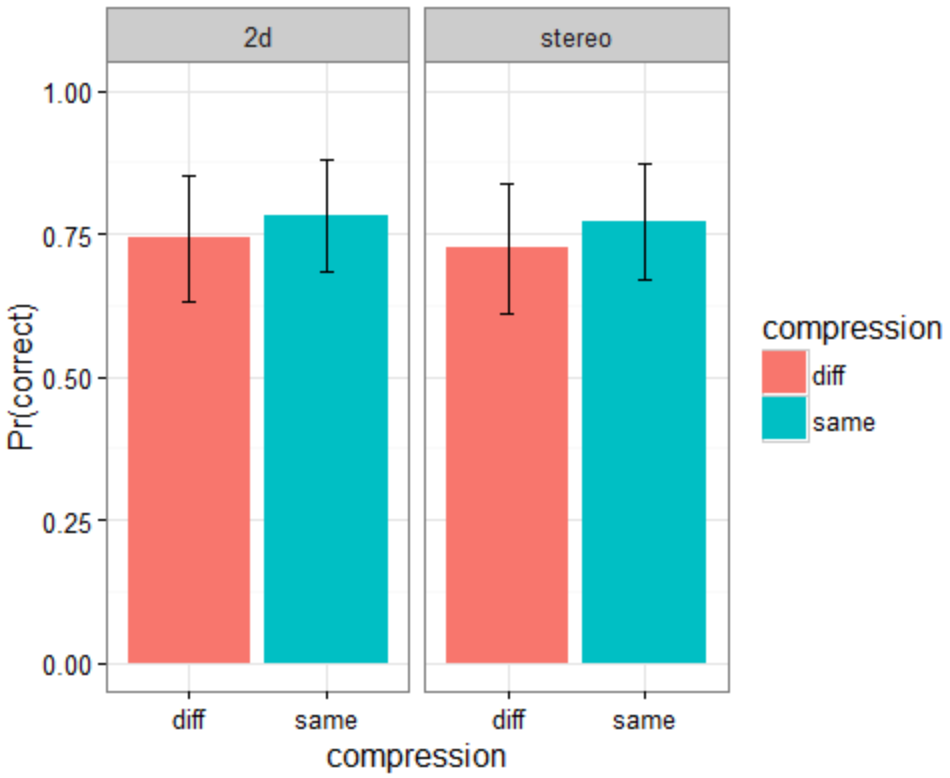



**Figure 4.7:** Different compression vs same compression for 2D and Stereo conditions averaged across observers and images. Error-bar represents ±1 Standard Error (SE)

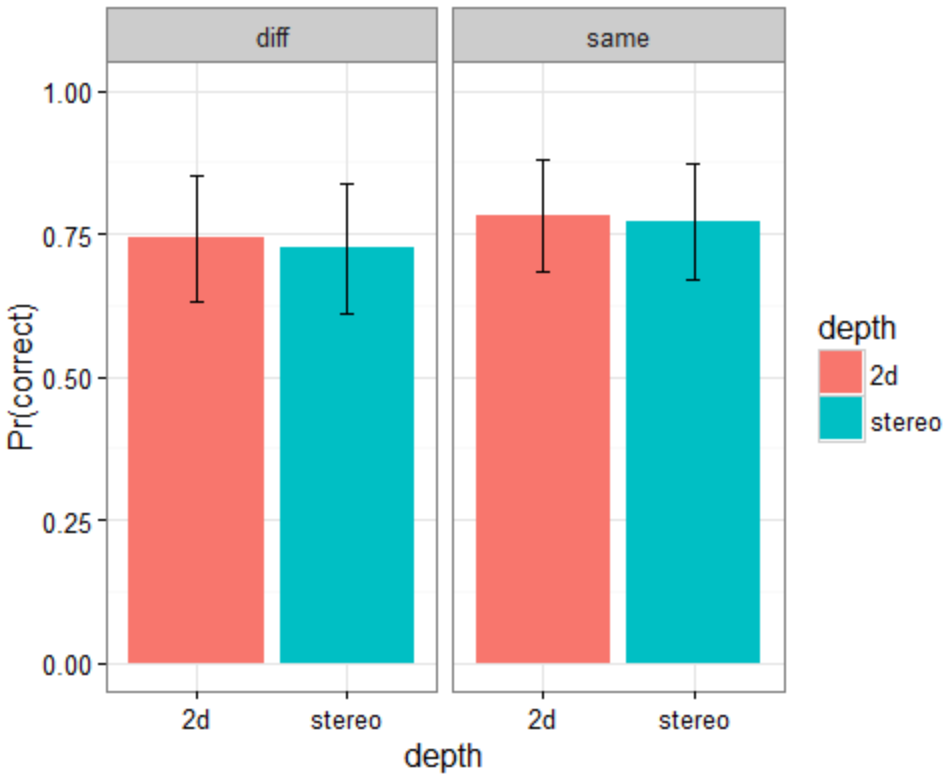



**Figure 4.8:** 2D vs stereo for same and different compression conditions averaged across observers and images. Error bars represents ±1 SE.

When treating the image as a random factor the interaction between compression and depth had a significant effect on artefact perception. Given that we need to understand this interaction before addressing our main hypotheses, I was interested to explore that interaction effect for each particular image. To see that I analysed the following model:

correct ~ ref * depth * compression + (1 | subject) (2)

where again the '*' operator implies the main effects as well as interaction effects. This formula takes image as a fixed effect instead of a random effect like Formula (1) with fixed effects of depth and compression and subject ID again acts as random effect to model repeated measures. The output of GLMM analysis using this model showed that for two specific images among the 12 images, Mandrill and Peacock, the depth-compression interaction had a significant effect on artefact perception. The Mandrill and Peacock images both have very fine texture which could be a reason behind this significant effect for depth-compression interaction.

**Table 2:** Effect of interaction between depth and compression for Mandrill and Peacock (Variables: 2 depth conditions (stereo, 2D), 2 compression conditions (same, different), 11 subjects, 2 reference images)

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| refMandrill.ppm:depthstereo: compressionsame | 1.33 | 0.40 | 3.41 | 0.0007 *** |
| refPeacock.ppm:depthstereo: compressionsame | 1.12 | 0.37 | 3.01 | 0.0026 ** |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | |

From the proportion correct data shown in Figure 4.9, I can say that for both images, artefacts were more visible with same compression than different compression when the images had disparity but in the zero-disparity condition (Figure 4.9) artefacts were more visible with different compression than same compression (opposite to the disparity case).

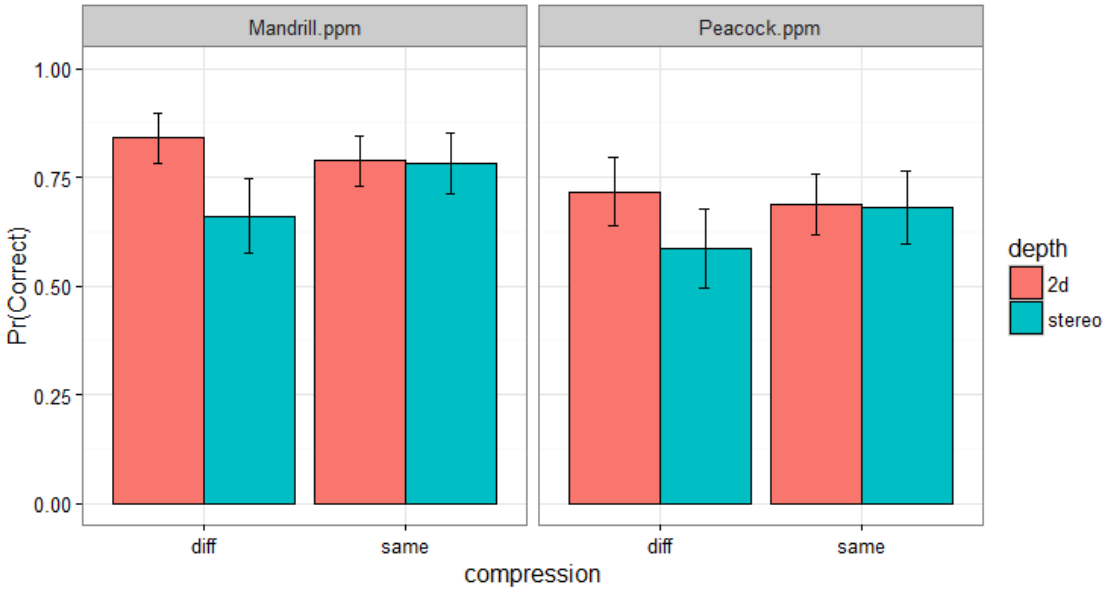



Figure 4.9: Same compression vs different compression for 2D and stereo in Mandrill and Peacock. Error bars represents ±1 SE.

From figure 4.5 we can see that the left (condition 3a) and right (condition 3b) same compression condition produced markedly different detection performance for Mandrill and Peacock, but not for other images. For both images, artefacts were more detectable in condition 3a (same compression left image) than condition 3b. The 2D same compression data shown in Figure 4.9 combines both condition 3a and b but the most appropriate comparison is with the left (3a) because that was used for the stereo same compression (condition 1). Figure 4.10 plots condition 3a (same_L) and condition 3b (same_R) separately. Note that when this data is used the same compression detections rate is larger or similar to the different compression case for both 2D and 3D conditions, consistent with the rest of the dataset.

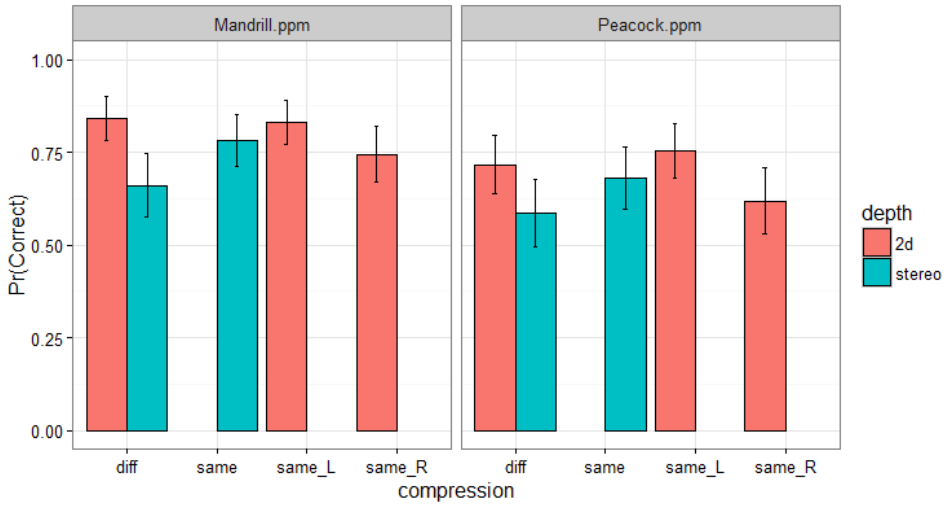



**Figure 4.10:** Comparison between left and right image for mandrill and peacock in 2D condition. Error bars represents ±1 SE.

**Hypothesis 1.1 Effect of Disparity:**

The above analysis shows that there is an interaction effect between depth and compression on artefact perception; however, this interaction seems to be mainly due to two images. Thus, while we need to be mindful of this interaction during the hypothesis testing we can consider our main hypotheses. To test hypothesis 1.1, I performed a series of planned comparisons between the proportion correct for the 2D and stereo conditions for each image (marginal means across compression level). This analysis was based on tests of linear contrasts estimating the difference between 2D and stereo predictions for each image using the lsmeans package in R to obtain the least square mean predictions. Least square means are means for groups that are adjusted for means of other factors in the model and are

less sensitive to missing data, they are better estimate of true population mean (SAS Work Shop 2018).

**Table 3:** Comparison between 2D versus Stereo for each image with DSC 1.2 compression.

| | Experiment1_hypothesis 1.1_2DvsStereo_two tail | | | | |
|---|---|---|---|---|---|
| contrast | ref | estimate | SE | p.value | Sig. |
| 2d - stereo | Barbara | 0.019 | 0.03 | 0.553 | FALSE |
| 2d - stereo | **CircularPattern26** | 0.064 | 0.03 | **0.042** | **TRUE** |
| 2d - stereo | **Clipboard** | -0.072 | 0.03 | **0.022** | **TRUE** |
| 2d - stereo | FemaleHorseFly | 0.033 | 0.02 | 0.077 | FALSE |
| 2d - stereo | HintergroundMusik | -0.035 | 0.03 | 0.270 | FALSE |
| 2d - stereo | **Landscape102** | 0.059 | 0.02 | **0.008** | **TRUE** |
| 2d - stereo | **Mandrill** | 0.069 | 0.03 | **0.013** | **TRUE** |
| 2d - stereo | MosaicBroadcom | 0.024 | 0.03 | 0.373 | FALSE |
| 2d - stereo | MysticMountain | -0.008 | 0.03 | 0.783 | FALSE |
| 2d - stereo | Noise | -0.007 | 0.01 | 0.517 | FALSE |
| 2d - stereo | Peacock | 0.044 | 0.03 | 0.149 | FALSE |
| 2d - stereo | **Tools** | 0.066 | 0.02 | **0.003** | **TRUE** |

The two-sided pairwise comparisons corresponding to hypothesis 1.1 are shown in Table 3 for each image. False Detection Rate (FDR) p-value correction was applied for the tests of these hypotheses at a significance level of 0.05. FDR adjustment determines adjusted p values for each test, it controls the expected proportion of false positive results in those tests (Benjamini and Hochberg 1995). Thus, it controls the false rejection of null hypothesis i.e., type I errors.

When the adjusted p value is less than the significance level ($p<0.05$), we can say that there is significant difference between the visibility of the artefacts (and hence the performance of the codec) for 2D and 3D images. In table 3, the rows with $p<0.05$ are bolded.

Table 3 shows that there was significant difference in relative performance of DSC 1.2 for CircularPattern26, Clipboard, Landscape102, Mandrill, Tools. Recall that Mandrill also demonstrated a significant interaction between depth and compression so the difference in table 3 needs to be interpreted with some caution (see Discussion). So, for these 5 images the performance of DSC 1.2 in 2D and the performance of DSC 1.2 in stereo was significantly different. For these 5 images, I cannot accept the null hypothesis $H_0$ (stereo-2D=0) and for remaining 7 images I failed to reject the null hypothesis as for those 7 images there was no significant difference for the performance of DSC 1.2 between 2D and stereo.

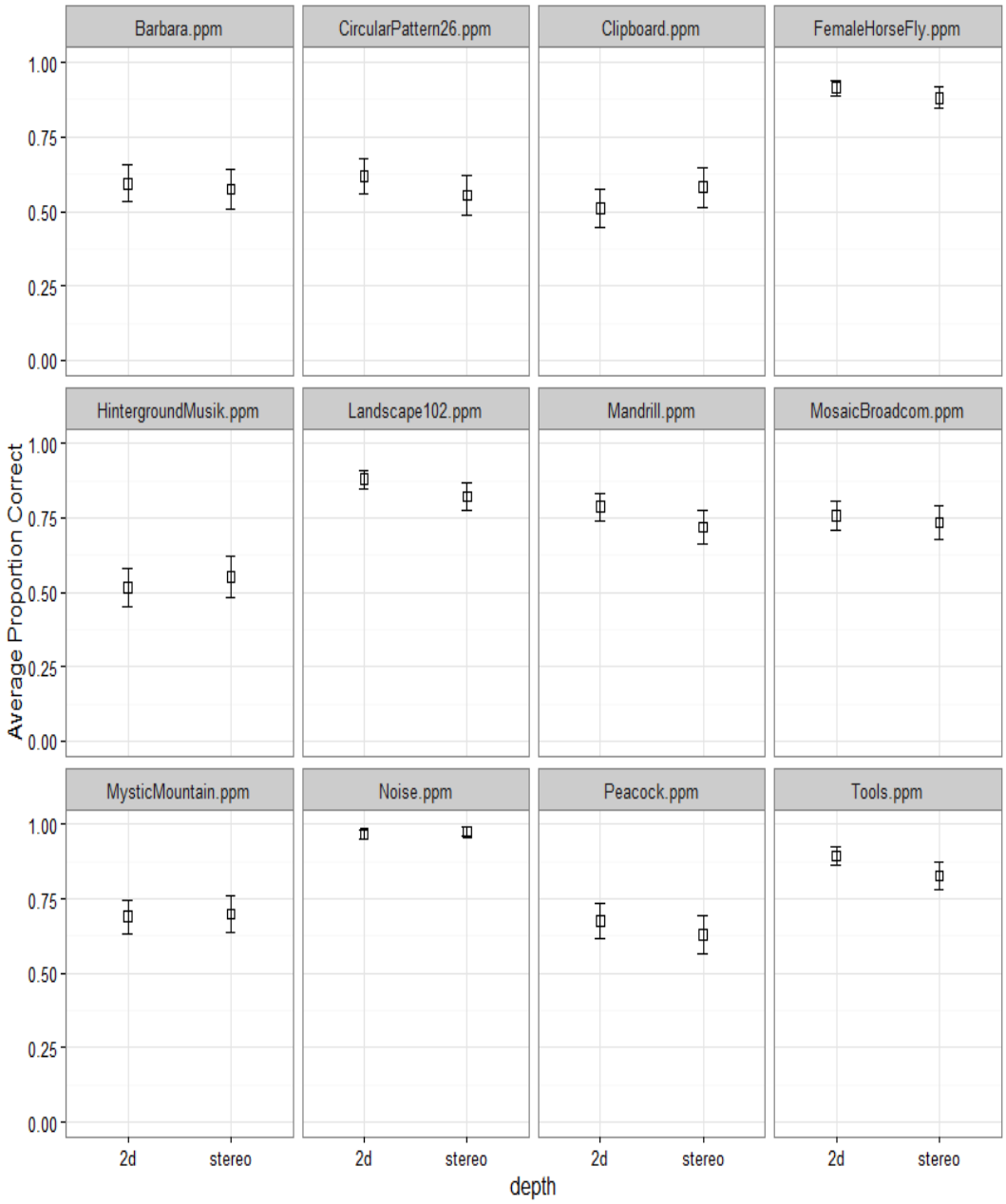



**Figure 4.11:** Average proportion correct for 2D versus stereo condition when DSC 1.2 compression was used (data for same & different compression are collapsed). Error bars represents ±1 SE.

We can visualize the difference between the performance of DSC 1.2 for 2D and stereo in the plot of Figure 4.11 which shows the average proportion correct for each image in the 2D and stereo conditions.

From the figure and difference estimates in Table 3 we can determine—for images where there is significant difference between 2D and stereo—in which condition (2D or stereo) did the DSC 1.2 codec perform better (artefacts less perceptible) or worse (artefacts more perceptible). The estimate values show that, among these 5 images, artefacts were more visible in the 2D condition than in the stereo condition for all images except Clipboard which means that for these images DSC 1.2 is more visually lossless in the stereo condition than the 2D condition.

**Hypothesis 1.2 Effect of Symmetry of Compression:**

In hypothesis 1.2, a comparison between different compression vs same compression was done.

**Table 4:** Comparison between different compression and same compression for each image with DSC 1.2 compression.

| | Experiment1_hypothesis 1.2_differentCompressionVSsameCompression_two tail | | | | |
|---|---|---|---|---|---|
| contrast | ref | estimate | SE | p.value | Sig. |
| diff – same | Barbara | -0.023 | 0.03 | 0.460 | FALSE |
| diff – same | CircularPattern26 | -0.020 | 0.03 | 0.527 | FALSE |
| diff – same | **Clipboard** | 0.068 | 0.03 | **0.033** | **TRUE** |
| diff – same | **FemaleHorseFly** | -0.056 | 0.02 | **0.004** | **TRUE** |
| diff – same | HintergroundMusik | 0.016 | 0.03 | 0.608 | FALSE |
| diff – same | **Landscape102** | -0.121 | 0.02 | **2.13e-07** | **TRUE** |
| diff – same | Mandrill | -0.014 | 0.03 | 0.603 | FALSE |
| diff – same | **MosaicBroadcom** | -0.099 | 0.03 | **3.12e-04** | **TRUE** |
| diff – same | MysticMountain | 0.029 | 0.03 | 0.325 | FALSE |
| diff – same | **Noise** | -0.028 | 0.01 | **0.006** | **TRUE** |
| diff – same | Peacock | 0.011 | 0.03 | 0.703 | FALSE |
| diff – same | **Tools** | -0.060 | 0.02 | **0.007** | **TRUE** |

Table 4 shows estimated differences and the p values; images with $p<0.05$ differences (two-tailed tests) are bolded indicating that the performance of DSC 1.2 was significantly different between the two compression symmetry types for the image. For Clipboard, FemaleHorseFly, Landscape102, MosaicBroadcom,

Noise and Tools the difference between the performance with different and same compression for DSC 1.2 was significant. For these images we cannot accept the null hypothesis $H_0$ (same-different=0). For remaining images, the difference was not significant, and we failed to reject the null hypothesis. Estimate values help us to determine, between different and same compression, in which condition artefacts were more perceptible. Table 4 shows that among 6 images with $p<0.05$, for all the images except Clipboard, artefacts were less visible with different compression than with same compression. The difference between the performance of different and same compression for DSC 1.2 is plotted in Figure 4.12 which shows the average proportion correct for each image in the two cases. Clipboard behaved differently than all other images in that artefacts were less visible with same compression than with different compression.

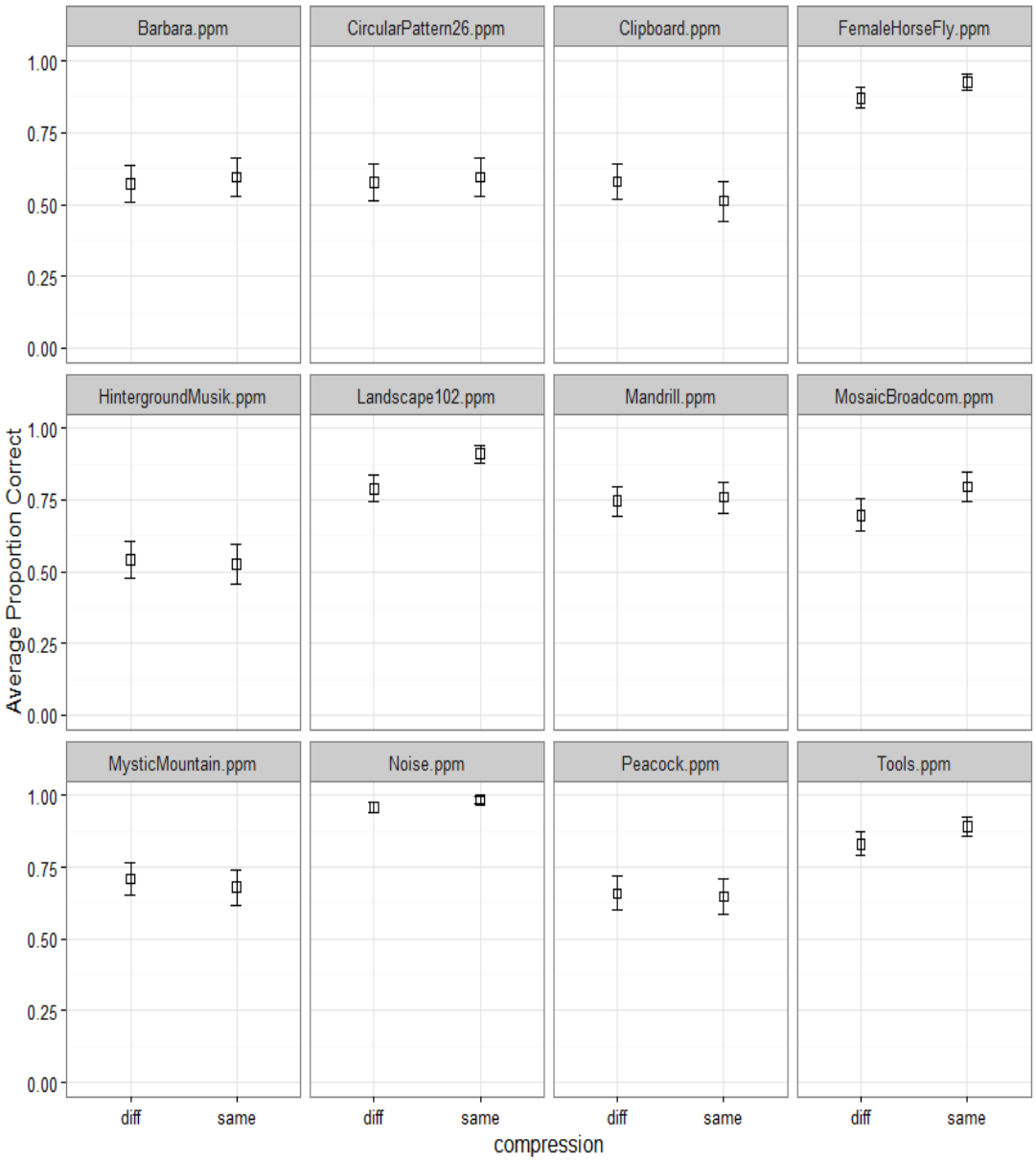



**Figure 4.12:** Average proportion correct for different compression versus same compression when DSC 1.2 compression was used. Error bars represents ±1 SE.

## 4.3 Discussion

In Experiment 1, we assessed the effects of (1) disparity of the images relative to the screen and (2) the matching or symmetry of the compression errors in the left and right images. Generally, I found that when there was a difference in the visibility of compression artefacts they were less visible in (1) 2D images presented with a disparity (‘stereo’ condition, offset from the screen) than in 2D images presented at the screen plane and (2) when the compression errors in the two eyes differed compared to when they matched.

I used 2D images (crops were different) that were previously used for extensive 2D testing of the DSC protocol. This previous testing can serve as a baseline for comparison for these results obtained in an S3D display. Specifically, the 2D images with same compression should be equivalent to the baseline tests of the DSC codec. Allison et al. (personal communication, 2018) did the subjective assessment of DSC 1.2 in 2D condition using the same crops I used in my experiment. The results of my experiment for 2D images with same compression were consistent with their results.

The use of 2D images with disparity allowed for assessment of the effects of displaying the images with a constant disparity with respect to the screen (a disparity pedestal) without the complications of variation in disparity within the

image as would be present in true stereoscopic images. For images where the 2D and 3D performance of the codec differed, the artefacts were more visible when the images were presented at the screen plane than with a disparity offset (stereo condition) except for the Clipboard image. In Section 2.4, I outlined several possible mechanisms that predict 2D and 3D differences. In the present experiment, the underlying mechanisms for this disparity pedestal effect cannot be due to stereoscopic factors such as spurious disparity or monocular features, distortion, rivalry and so on as the images in the two eyes were identical in the same compression case, except for an offset. Many perceptual judgements are most precise when the stimulus is foveated by both eyes and performance degrades with a disparity pedestal or fixation disparity (Regan and Gray 2009; Zaroff, Knutelska, and Frumkes 2003). However, this is unlikely to be the explanation for the pedestal effect in the present experiment as the subject's fixation was not controlled and they were able and likely to converge on the stimulus during testing. This convergence though could indirectly be the cause of the disparity effect, the vergence demand but not the accommodation demand for the 2D and 3D conditions differed. Stereoscopically, the 2D images were on the screen plane while the 3D images were behind the screen plane but in both conditions optimal focus was at the screen distance. Normally when converging at a further distance the viewer should also accommodate at the further distance. If the observer's accommodation shifted away from the screen plane when

converging on the disparate stimulus (vergence accommodation) (Maddox 1886), this would result in more defocus blur for the 3D image than for the 2D case (Hoffman et al. 2008; S. Yang and Sheedy 2011). Mismatch between vergence and accommodation can also increase difficulty in fusing the stimuli and produce fixation disparity. This might have reduced the users' artefact detection ability for 3D images compared to 2D. However, the depth offset used in this experiment was modest and the predicted blur if the eyes focused on the far target would only be about 0.1 D, within the normal depth of field of the eye (Marcos, Moreno, and Navarro 1999), so any blur effect would be small.

The use of 2D images also allowed for assessment of the effects of compression artefacts when they were identical in the two eyes' images compared to when they differed, Generally, DSC 1.2 was more visually lossless when different compression was applied compared to when same compression was applied. Binocular summation could be a reason for perceiving artefacts more in same compression than different compression. Due to binocular summation the detection threshold for a stimulus is lower with two eyes than one eye. Visual acuity, contrast sensitivity, flicker perception, and brightness perception (Howard and Rogers 2012) can be improved by binocular summation as the information received in each eye are combined. Dichoptic masking could be a reason for perceiving artefacts less in the different compression. Dichoptic masking occurs when both eyes see the images of similar pattern. The detection of the test stimulus

in the one eye is prevented by the ‘mask’ stimulus in another eye. Masking is substantial when the images are similar.

For two specific images among the 12 images, Mandrill and Peacock, the depth-compression interaction had a significant effect on artefact perception. For both images, artefacts were more visible with same compression than different compression when the images had disparity (consistent with my general findings) but in the zero-disparity condition (Figure 4.9) artefacts were more visible with different compression. The zero-disparity same compression case consisted of two sub-conditions, one where the compressed image corresponded to the left image of the zero-disparity different compression case (condition 3a) and the other corresponded to the right image of the zero-disparity different compression case (condition 3b). In most cases the performance on these two cases was very similar but for the Mandrill and Peacock images the artefacts were noticeably more visible in the left image (Figure 4.5). The Mandrill and Peacock images both had very fine texture which could be a reason behind the significant differences in artefacts in conditions 3a and 3b. Figure 4.13 and Figure 4.14 point out the locations in the images (with disparity) where the difference between different compression (DC) and same compression (SC) are visible. These figures were obtained using S-CIELAB ( Poirson and Wandell 1997), the results in errorImage (difference between two images) were calculated in CIELAB delta E units and the figures show the spatial distribution of the errors, highlighting errors that are 10 delta E units or

larger. Since these images have disparity, the common artefacts in both the left and right images appear twice with an offset corresponding to the disparity (so all artefacts are repeated except at the edges in the SC case). We can see that in the DC case that for the indicated locations (enclosed in boxes) the artefacts are not repeated or are attenuated in one copy indicating more artefacts (noise) in one image than in the other. For these two images the artefacts are more apparent in the left than the right eye consistent with the difference in detection of these artefacts in conditions 3a and 3b. The stereo same compression case used the same compressed source images as condition 3b. We showed in Figure 4.10 that when this data was used (instead of the combined condition 3a and 3b data) then both the stereo and 2D results for these images conformed to the general pattern of results where same compression was more visible than different compression.

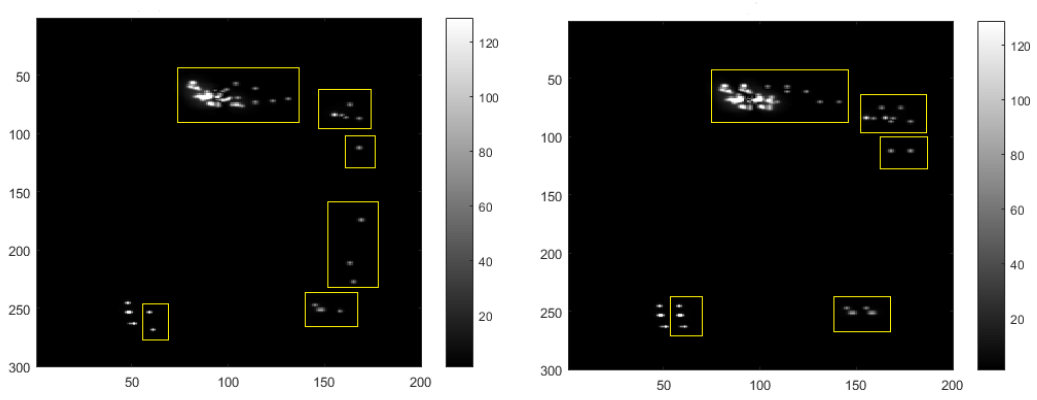



**Figure 4.13:** Difference between Left and Right images for Mandrill (different compression (left one) vs same compression (right one) in 3D) using S-CIELAB (showing spatial distribution of the errors)

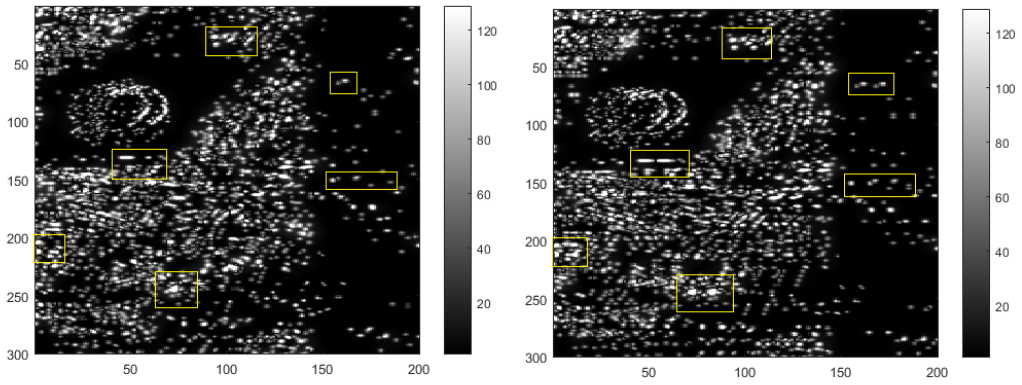



**Figure 4.14:** Difference between Left and Right image for Peacock (different compression (left one) vs same compression (right one) in 3D)

The exception to both the compression and disparity patterns was the Clipboard image where, artefacts were more perceptible for different compression than same compression and more visible in the stereo condition than in the 2D condition. The compression noise generated in the two eyes' images of clipboard may be uncorrelated or independent which could make the artefacts more apparent in the 3D condition compared to 2D. Figure 4.16 shows differences in compression artefacts in the Clipboard images for the case where the images are in correspondence (with zero disparity). Thus, in the SC case when the left and right images are subtracted the images are identical and there is no difference in the resulting image. The DC images show locations where the artefacts do not match in the two images. It can be seen that there are significant regions where the artefacts do not match in the two eyes. Figure 4.17 shows the S-CIElab errors in

the left and right images (with disparity) relative to the reference. The errors appear similarly distributed as in the L-R difference image and most of the errors in each eye appear to be reflected in the difference errors. The strength of the difference signal in the different compression case (i.e. most of the artefacts do not match) may explain why artefacts were more visible in this condition. Significant regions of artefact are introduced along the edge of the image when shifted and this may underlie the increased detection rate in the 3D case.

The most obvious difference between Clipboard and other images is that Clipboard contains text (Figure 4.15) while there is no text in any other image. However, there appears to be little compression noise near the text in Figure 4.16 suggesting this is not the cause of the unusual pattern of results with the Clipboard image. Note that performance for Clipboard is near chance. Similarly, several of the images that showed non-significant effects of disparity or compression type had performance near chance (Barbara, CircularPattern and HintergroundMusik). Conversely, performance for noise was near ceiling. It is difficult to compare between conditions for images where performance is either near chance or ceiling as the statistical power of the comparisons is low (Wichmann and Hill 2001; McKee, Klein, and Teller 1985).


**Figure 4.15:** Clipboard

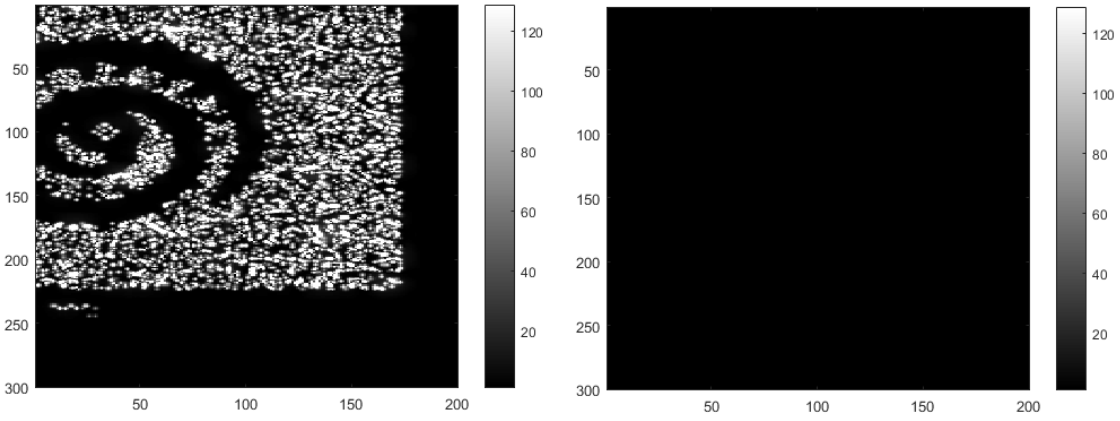



**Figure 4.16:** Difference between Left and Right image for Clipboard (different compression (left one) vs same compression (right one) in 2D)

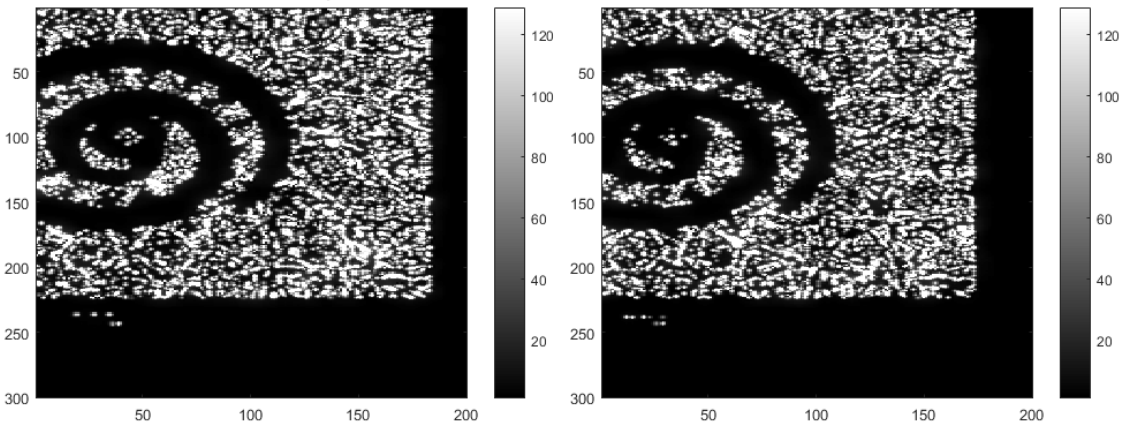



**Figure 4.17:** Difference between uncompressed and compressed image for Clipboard (Left eye (left one) vs right eye (right one))

# Chapter 5
# Experiment 2: 2D vs 3D using stereo images

In this experiment I used stereoscopic images to asses the effect of disparity of the Images or detectability of compression artefacts. The goals of the experiment were to see the performance of the codec in 2D vs 3D condition and to compare between two different level of compression. This allowed for assessment of the effects of disparity and to investigate whether there were images for which the performance of VDC-M 1.0.7 was different between disparity and no disparity conditions.

## 5.1 Subjects and Conditions

In Experiment 2, eleven subjects from the York University Community participated. All passed the vision screening tests, and no one was excluded for failing to detect control images. Participants ranged in age from 21 to 39. Among the participants four were female and seven were male. Each participant was paid $20/hour.

In this experiment I used stereoscopic images and the VDC-M 1.0.7 was tested under three display conditions at each of two different levels of compression, 4:1 compression (6bpp) and 6:1 compression (4bpp). The display conditions were as follows:

1. Stereoscopic image pairs with disparity between left eye and right eye image.

2 a. Left image presented to both eyes. In this case there is no disparity

between the left eye and right eye images.

2 b. As in condition 2a except the right image is used.

The first condition was the normal stereoscopic condition and the second and third condition were 2D conditions for comparison. Like experiment 1, in this experiment I also presented a cropped portion (600 X 500 pixel) of an image instead of the full-size image. The source images were 24-bit RGB stereoscopic images that I resized to 1920 x 1200 pixels to produce the reference images.

For condition 1, subjects were presented separate left and right views of the same scene as stereoscopic images. So, there was disparity between images and they appeared in stereoscopic 3D depth behind the screen plane. Cropping was done after compression producing three crops for each stereo pair of images: one pair at 4:1 compression, another pair at 6:1 compression and one pair uncompressed. For condition 2 a and 2 b, only the cropped left image or cropped right image of the pair were used, respectively.

VDCM 1.0.7 compressions at both compression levels were tested for all the three conditions by all eleven viewers. That is, each viewer tested VDCM 1.0.7 4:1 compression for all the viewing condition and image combinations and tested VDCM 1.0.7 6:1 compression for all the viewing condition and image combinations.

## 5.2 Stimuli

In experiment 2, twelve stereoscopic images were used which were selected from 1003 images. Selection was based on 4:1 compression but the same set of images was used for both 4:1 and 6:1 compression.

I collected 980 stereoscopic images from the image sharing website Flickr (Flickr website 2018) and 23 images were selected from the Middleburry dataset which is used as a ground truth dataset for evaluating computer vision algorithms (Scharstein et al. 2014) . In the Flickr source images left and right eye images were combined side-by-side in one image so were separated into individual left and right eye images. All the 1003 images collected were in jpg format and I converted them to ppm format for lossless intermediate processing. As the source images were in jpg format they are already compressed, this compression may affect the artefact detection for VDC-M compression. But as jpg is one of the most commonly used formats for images, it is also desirable to see the effects of additional compression artefacts of VDC-M on the perceived quality of jpeg-formatted source images. As all the Flickr source and Middleburry images were larger than 1920 x 1200 in pixel resolution, I resized (by using ‘bicubic’ interpolation which is the default method for ‘imresize’ function in MATLAB) the Flickr and Middleburry source images to monitor resolution and then flipped the images so that in the mirror stereoscopic display viewers could see the actual image instead of its reflected version. Then these images were compressed by the VDC-M 1.0.7 codec at 4:1 compression

level. These were subsequently decompressed to produce 24-bit images for display corresponding to the compressed images. PSNR for the set of compressed images relative to the Flickr source images was computed and sorted based on following criteria.

- The 30 lowest PSNR in either the left or right eye image (i.e. if the left eye had the lowest PSNR then the left, if the right eye had the lowest PSNR then we would use that). The goal was to pick the lowest PSNR images regardless of which eye. (these were the most difficult images for the codec i.e. these images should be lossy as their PSNR was lowest).
- The next 10 images from the remaining 950 images with the lowest average PSNR of left and right eye image. This selected for images with low PSNR in both eyes' images.
- The next 10 images from the remaining 940 images with the biggest difference of PSNR between left and right eye image. This selected for images where the compression artefacts were unbalanced and predominantly in one half image.

Then I viewed these 73 images in the stereoscopic display and discarded images with excessive disparity or mismatch in focus, contrast or exposure between the two images (i.e., images that caused excessive visual discomfort (Lambooij et al.

2009; Tam et al. 2011) ),or had poor colour or excessive blur. Based on these criteria I selected 31 images (13 image from Flickr and 18 images from the Middleburry set) for the next step. Then three experienced observers identified perceptible artefact regions in the images using the flicker paradigm on the entire (not cropped) images. Forty-five artefact or flicker locations were selected from thirty-one images: twenty-four crops from thirteen Flickr source images and the rest of the crops were from the 18 Middleburry images. The crop size was 600 x 500 pixels. I viewed the reference and compressed cropped images in the stereoscope and some had excessive (diplopic) disparity or window violations that created visual discomfort for these image pairs in the stereoscope.

Additionally, the Middlebury images were taken with parallel cameras and always appeared in front of the screen. To improve visual comfort the resized and flipped images were shifted and the convergence was adjusted: the left images were shifted to right and right images were shifted to left to change the apparent position of the objects relative to the screen until a comfortable view was obtained. Then I cropped the shifted image in the original crop position. A comfortable view could not be obtained for four image crops and these were excluded leaving 41 cropped images.

Based on flicker testing with these images I selected 23 cropped images based on following criteria:

- Artefacts were visible in both 2D and stereoscopic view
- Artefacts were difficult to find out in both 2D and stereoscopic view
- Artefacts were more visible in 2D compared to S3D view

Among these 23 selected images, based on criteria 1, I selected 9 images, 9 images were selected according to 2$^{nd}$ criteria and 5 images were included based on 3$^{rd}$ criteria. The final twelve crops were selected from these 23 cropped images based on the opinion/perception of two expert viewers. In this way I finalized a set of twelve images shown in Figure 5.1, Fig. 5.2 and Fig. 5.3.


**Figure 5.1:** Four cropped images for experiment 2

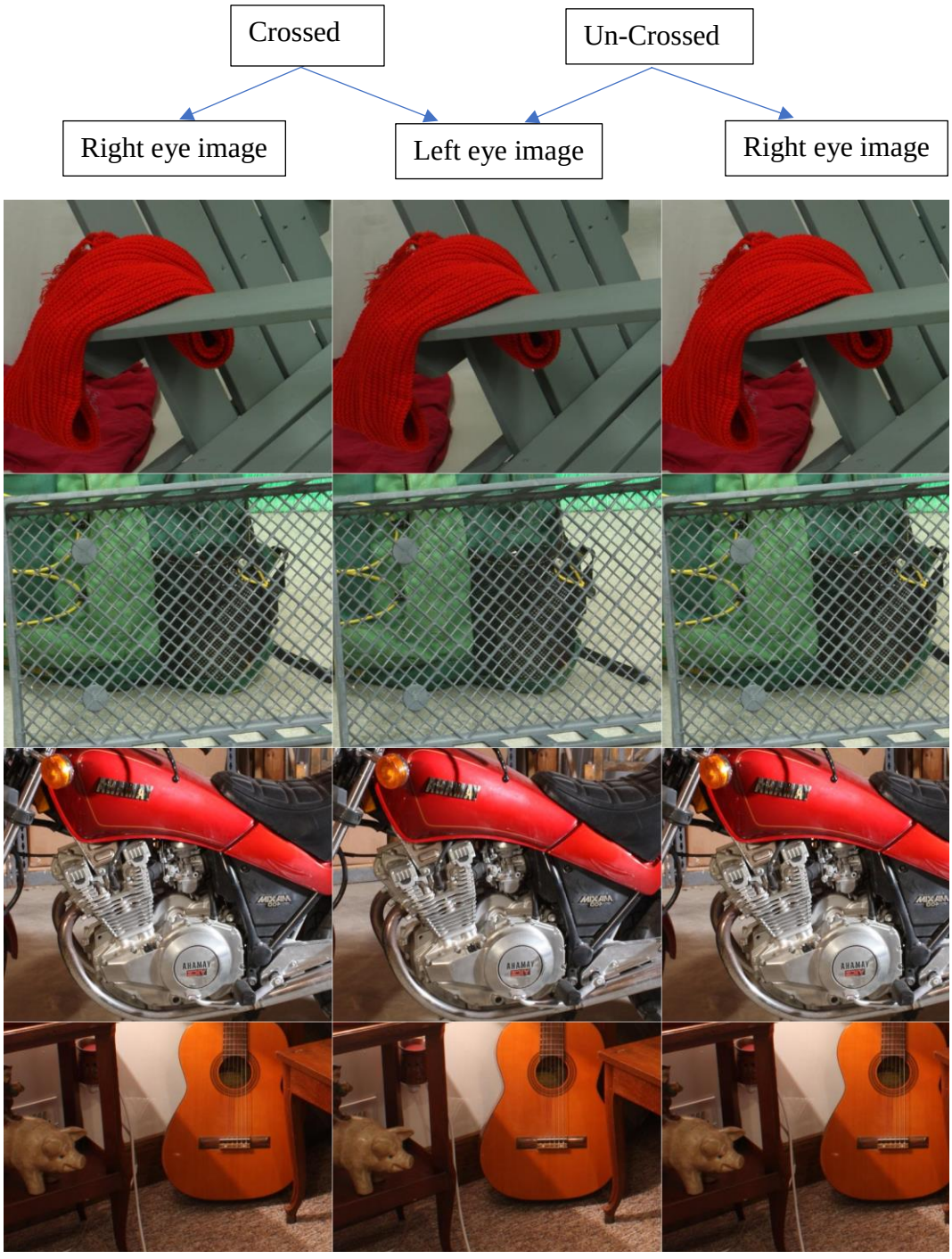



**Figure 5.2:** Four cropped images for experiment 2

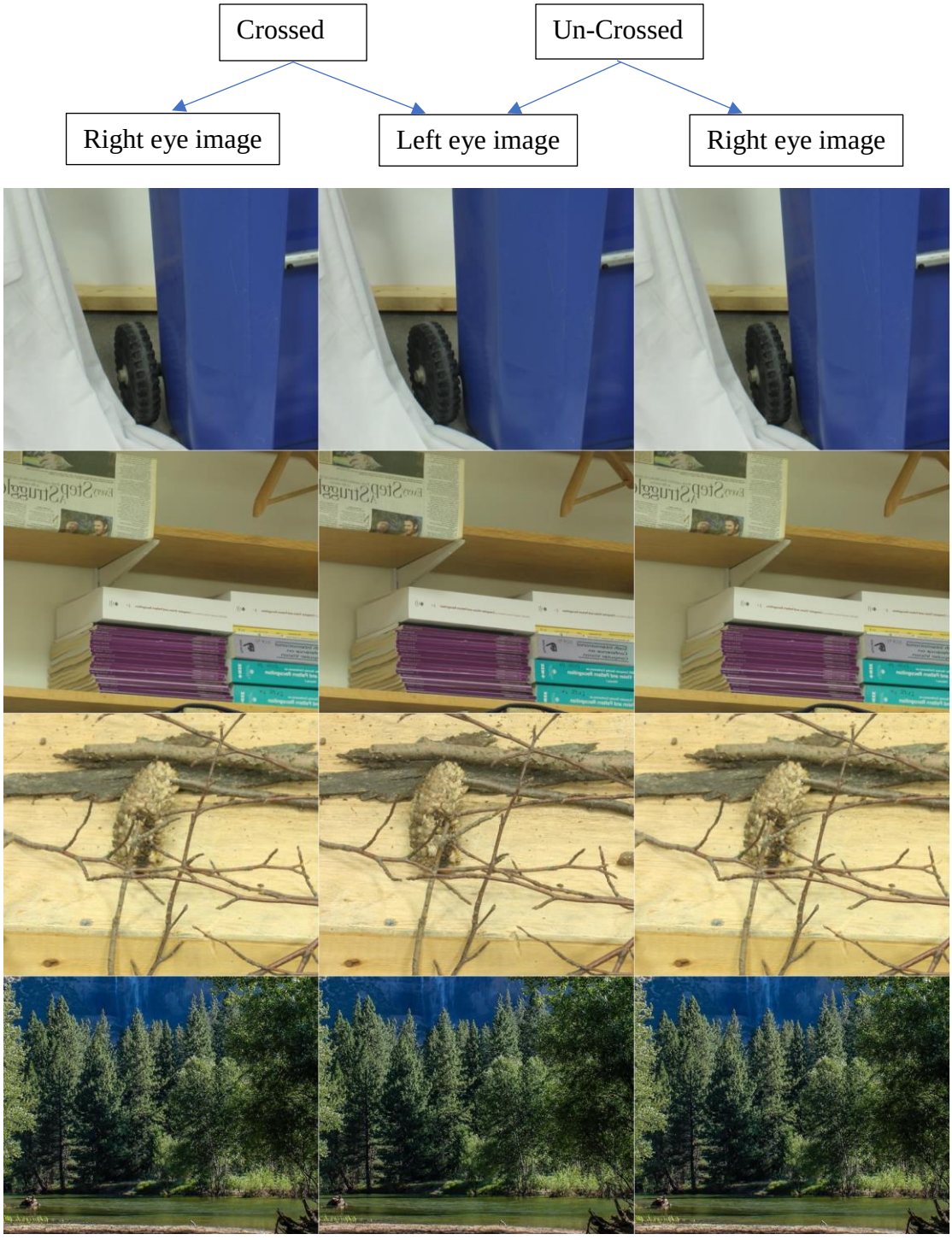



**Figure 5.3:** Four cropped images for experiment 2

The images and crops were selected based on analysis and viewing at 4:1 compression but the same choices were used for both 4:1 and 6:1 compression. An identical process of compression, shifting and cropping was performed on the same set of twelve images with VDC-M 1.0.7 at 6:1 compression. This resulted in two sets of 12 images with two different level of compression.

## 5.3 Procedure

The codec was tested for two of the three conditions at each of two different levels of compression for a given viewer. Six subjects were tested for condition 1 and 2a and remaining 5 subjects were tested for condition 1 and condition 2b. I did not expect systematic differences between 2D viewing of the left versus right image, but this procedure allowed me to evaluate this assumption. For each condition all 12 images were presented to the user. For each level of compression, the trials were divided into 10 blocks containing two trials per image per condition, presented in random order. So, each of the 12 images was displayed 20 times for each condition resulting in 240 experimental trials per condition at each compression level and each of the 3 control images was displayed 20 times for each compression level resulting in 60 control trials per compression level. While the control images were not compressed with the target codec, they were included in sessions for both codecs. Thus, the experiment consisted of 1080 trials per observer ((240 experimental trials x 2 conditions x 2 compression levels) + (60

control trials x 2 compression levels)). The control images were heavily compressed using JPEG 2000 with a quality 10, the purpose of using these images was the same as in experiment 1.

As the cropped image size was larger than in experiment 1, the viewing time needed to be increased. I set the viewing time to 8 s based on pilot testing that suggested this gave the subject sufficient time to scan the crop and find areas that flickered. Total time calculated for 20 blocks plus the time for vision screening was 3 h 20 minutes (1 h 40 minutes for each compression level). As this is a long time for a participant to keep their concentration, I divided each experiment (one with 6:1 compression level and another 4:1 compression level) into two sessions. The first session was about 1hr, allowing for vision screening and running 5 blocks. The 2$^{nd}$ session was about 40 minutes when the rest of the 5 blocks were tested. I maintained a minimum 1 h gap between the two sessions so that the users could rest. The task of participants in this experiment was same as in the previous experiment. The next image pair appeared after the participant gave his/her response.

## 5.4 Results

In experiment 2, VDC-M 4:1(6 bpp) and VDC-M 6:1(4 bpp) were tested for different viewing conditions. I tested 4 hypotheses:

**Hypothesis 2.1:** As in experiment 1, I compared the performance of the compression algorithm for stereoscopic vs 2D conditions regardless of compression rate. The main hypothesis to be tested was that the sensitivity to VDC-M compression artefacts differed between stereoscopic images (condition 1) and equivalent binocularly-viewed 2D images (conditions 2a and 2b). Thus, testing of Hypothesis 2.1 assessed whether there was a significant difference between compressed images with **disparity[stereo]** and compressed images with **no disparity[2D]** (i.e., the null hypothesis was that stereo-2D = 0).

**Hypothesis 2.2:** As floor effects are less likely for 6:1 compression the effect of viewing condition might be more apparent here. Therefore, I hypothesized that performance in the stereo vs 2D conditions might differ, but for only 6:1 compression of VDC-M 1.0.7. Thus, testing of Hypothesis 2.2 was based on linear contrasts comparing the stereo vs. 2D conditions for the for 6:1 compression case.

**Hypotheses 2.3 and 2.4:** I tested VDC-M for two compression rates: 6 bpp and 4 bpp; I expected that the compression errors and hence the compression performance at 4 bpp should be worse (artefacts more detectable) than the performance at 6bpp for both stereo and 2D conditions. Tests of Hypothesis 2.3 assessed whether artefacts were more detectable for compressed **stereoscopic**

images at **4 bpp** compression than at **6 bpp** (i.e., the null hypothesis was that 4bpp-6bbp <= 0 for stereo images). Tests of Hypothesis 2.4 assessed the equivalent comparison for **2D** images (i.e., the null hypothesis was that 4bpp-6bpp <= 0 for 2D images).

For above all hypothesis tests we compared the relative performance of VDC-M between different conditions. Our aim was not to find out for which condition VDC-M was visually lossless or not. All the data collected were fitted using GLMM to test the above hypothesies.

**Descriptive Result:**

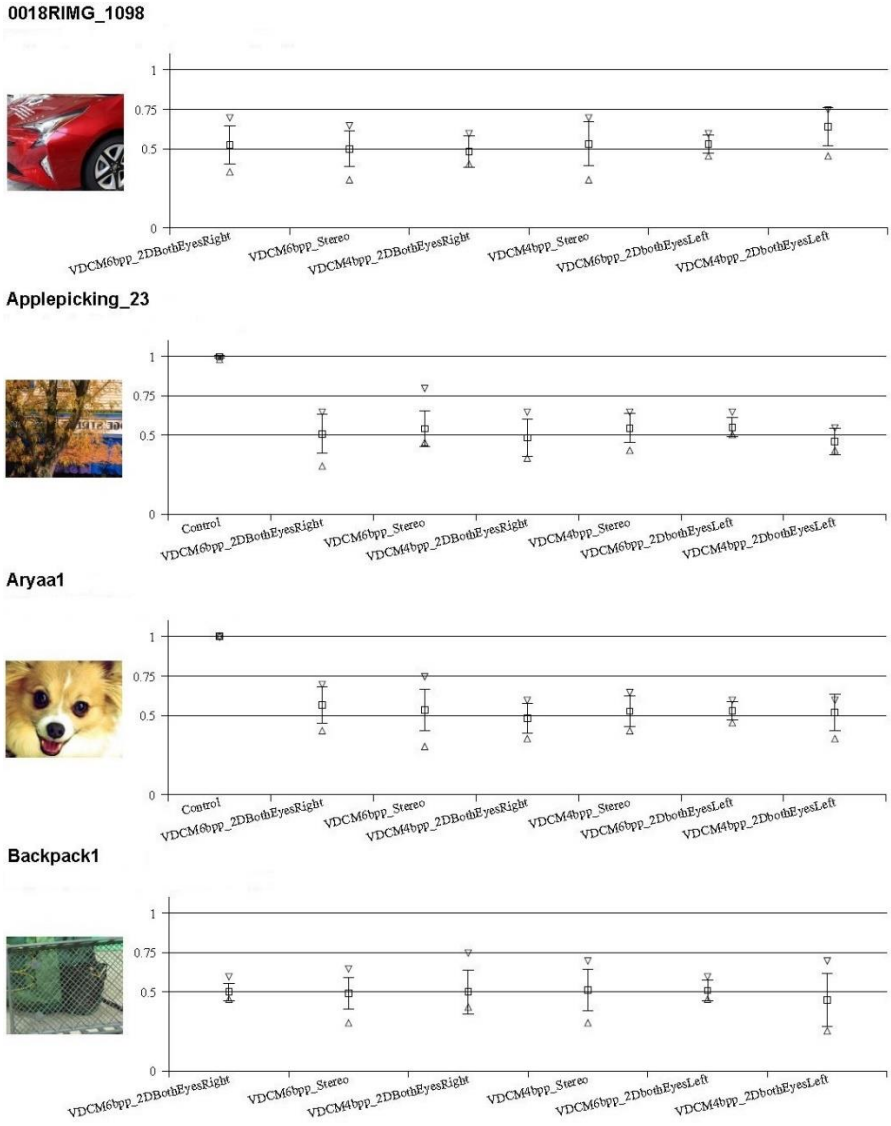



**Figure 5.4:** Sample proportion correct under different conditions for VDC-M 1.0.7 (6bpp and 4bpp) compression (part 1) (6bpp = 4:1 compression, 4bpp = 6:1 compression, 2DBothEyesRight = Right image in both eyes, 2DBothEyesLeft = Left image in both eyes). Error bars represents ±1 SD.

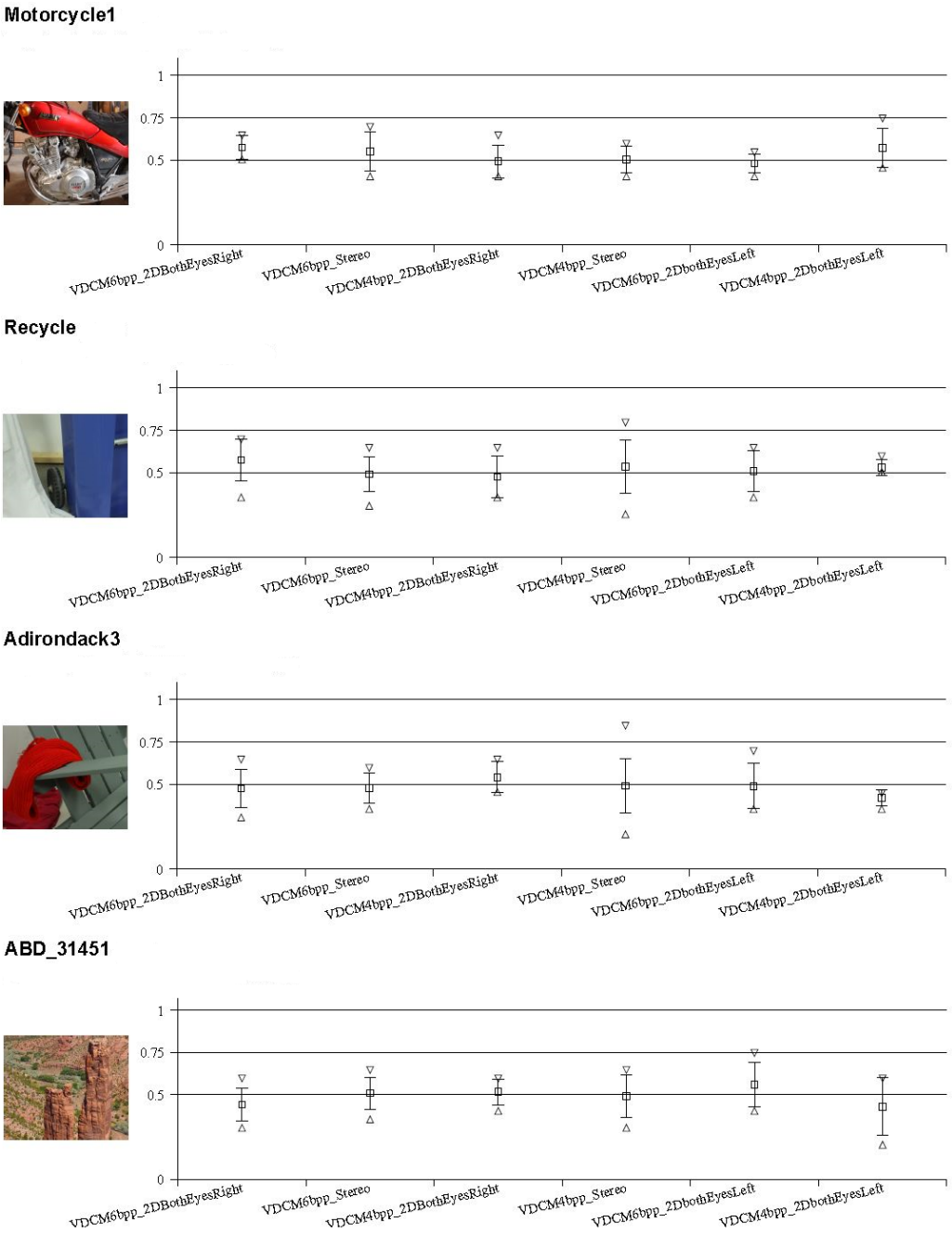



**Figure 5.5:** Sample proportion correct under different conditions for VDC-M 1.0.7 (6bpp and 4bpp) compression (part 2). Error bars represents ±1 SD.

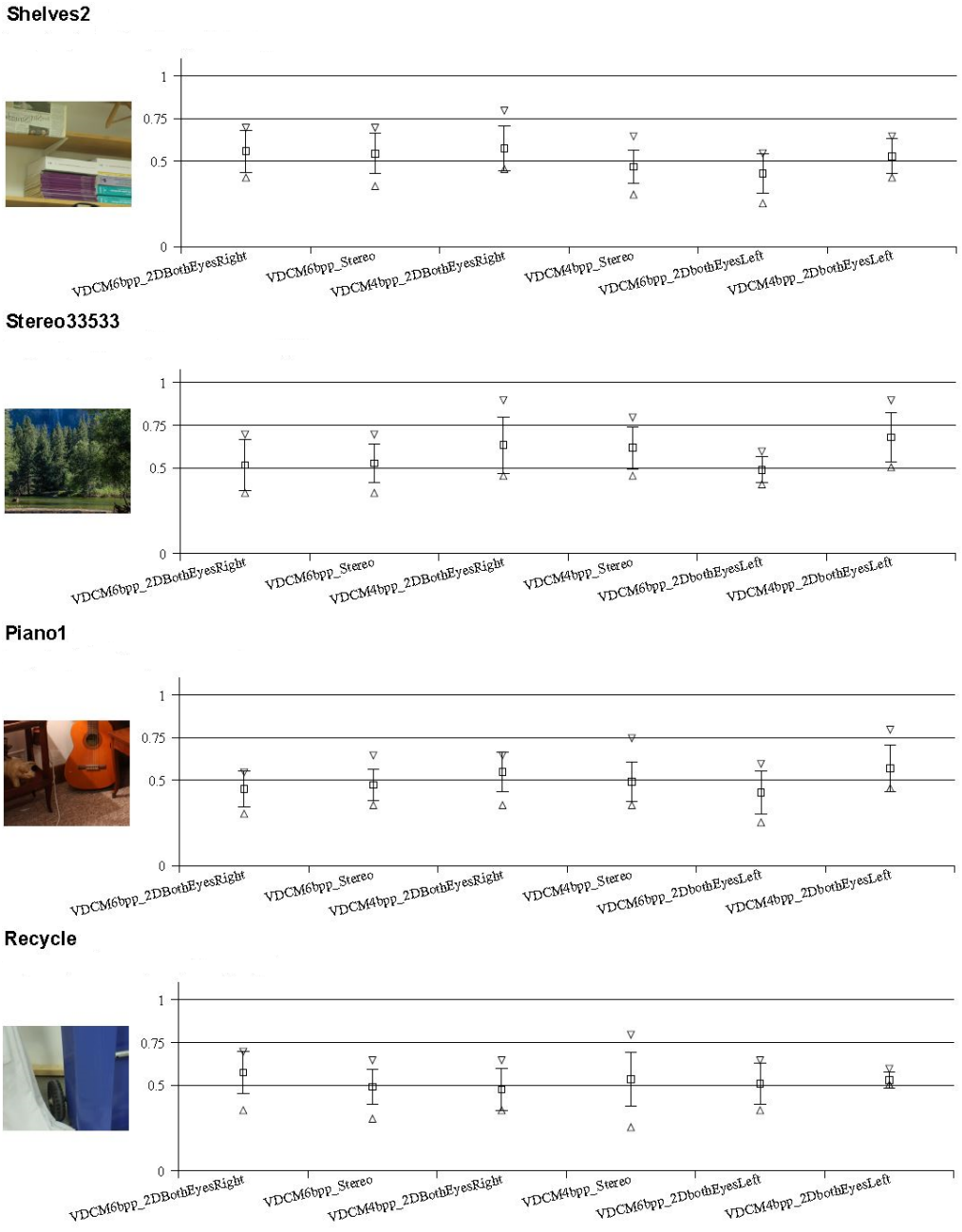



**Figure 5.6:** Sample proportion correct under different conditions for VDC-M 1.0.7 (6bpp and 4bpp) compression (part 3). Error bars represents ±1 SD.

Fig. 5.4, 5.5 and 5.6 shows proportion correct score data for VDC-M 1.0.7 (both 6: 1 and 4:1 compression) for each image tested for Experiment 2. In the figure, the mean proportion correct across observers is represented by a square, error bars represents ±1 standard deviation and the range of scores is indicated by upward and downward triangles. All data was included as all subjects had proportion correct above 0.95 for control images.

This figure helps us to get an idea in which conditions VDC-M 1.0.7 was visually lossless or not. The codec meets the visually lossless criteria (based on the mean+standard deviation version) in all conditions except in the Stereo33533 image at 4 bpp. although some observers performed above 0.75 in a few of the 4 bpp condtions. It also gives us a rough idea of relative performance, but we cannot make pairwise within subjects comparisons through this plot.

**GLMM analysis:**

Like the analysis for Experiment 1, I did a GLMM analysis for this experiment. The R statistical software environment (R Core Team 2017) was used for the analysis and data was fitted through maximum likelihood estimation using the 'glmer' procedure in package lme4. Before fitting GLMM, control data was excluded because it did not include factors of interest and including control data can cause convergence failure as the response proportions were all 1.0 for control data. My main hypotheses were that there would be significant effects of both depth and

compression ratio. Depth has two levels: stereoscopic (stereo, Conditions 1) and 2D (Conditions 2a,b); whereas bpp (bit per pixel) also has two levels: 6 bpp (4:1 compression level) and 4 bpp (6:1 compression level). To address the hypotheses and to test for possible interaction of depth and bpp on the artefact visibility, I used the following formula for the model (see section 4.2.2 for notation):

correct ~ depth * bpp + (1 | subject) + (1 | ref)

For the above formula, the depth and bpp are modeled as fixed effects whereas subject ID and image (ref) are considered as random effects. The error distribution of the response variable was modelled as a binomial distribution.

The summary of the GLMM analysis for the above formula is as following:

**Table 5:** Effect of interaction between depth and compression for 2nd set of images (Variables: 2 depth conditions (stereo, 2D), 2 compression (bpp) conditions (6, 4), 11 subjects, 12 reference images)

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 0.096 | 0.05 | 2.05 | 0.041 * |
| depthstereo | -0.006 | 0.06 | -0.11 | 0.912 |
| bpp6 | -0.057 | 0.06 | -1.05 | 0.295 |
| depthstereo:bpp6 | 0.023 | 0.08 | 0.30 | 0.770 |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | |

The above summary shows that there was no significant effect for interaction between depth and bpp for random image and random subject. There were also no significant main effects.

Despite the lack of main effects, I had specific hypothesises which I followed up with two further analyses. The model formulas for assessing the hypotheses are as follows:

correct ~ ref * depth + (1|subject)

correct ~ ref * bpp + (1|subject)

where 1st formula is for testing hypothesis 2.1 and 2.2 and the 2nd formula is for testing hypothesis 2.3 and 2.4. The model assumed that image (ref) was a fixed effect for these analyses so that I could estimate the effects of depth and bpp on individual images. For hypothesis 2.1 and hypothesis 2.2, two-sided pairwise comparisons were done while one-sided pairwise comparisons were used for hypothesis 2.3 and hypothesis 2.4 as we predict higher compression ratios produce more artefacts. Significance level for all hypotheses was set at 0.05 and False Detection Rate (FDR) p-value correction was applied in all hypothesis tests. To test hypothesis 2.1, I compared proportion correct data between no disparity (2D) and disparity (stereo) for both 4:1 and 6:1 compression rate (i.e., collapsed across compression levels). From the results of these comparisons presented in Table 6, I can see that there were no significant differences between 2D vs stereo for any of the 12 images. The data are plotted in Fig. 5.7 which shows the mean for each image in the 2D and stereo conditions and confirms the similarity of the 2D and stereo performance. Thus, I can conclude that performance of VDC-M 1.0.7 with stereo images was not significantly different from the performance of VDC-M 1.0.7 with 2D images.

**Table 6:** Comparison between 2D versus Stereo for each image with VDC-M 1.0.7 (6:1 & 4:1) compression.

| | Experiment2_hypothesis 2.1_2DvsStereo_two tail | | | | |
|---|---|---|---|---|---|
| contrast | ref | estimate | SE | p.value | sig |
| 2d – stereo | 0018RIMG_1098 | 0.025 | 0.03 | 0.457 | FALSE |
| 2d – stereo | ABD_31451 | -0.014 | 0.03 | 0.686 | FALSE |
| 2d – stereo | Adirondack3 | 2.51e-06 | 0.03 | 0.999 | FALSE |
| 2d – stereo | Applepicking_23 | -0.043 | 0.03 | 0.199 | FALSE |
| 2d – stereo | Aryaa1 | -0.007 | 0.03 | 0.839 | FALSE |
| 2d – stereo | Backpack1 | -0.111 | 0.03 | 0.736 | FALSE |
| 2d – stereo | Motorcycle1 | 0.002 | 0.03 | 0.946 | FALSE |
| 2d – stereo | Piano1 | 0.018 | 0.03 | 0.589 | FALSE |
| 2d – stereo | Recycle | 0.009 | 0.03 | 0.787 | FALSE |
| 2d – stereo | Shelves2 | 0.021 | 0.03 | 0.543 | FALSE |
| 2d – stereo | Stereo33533 | 0.007 | 0.03 | 0.838 | FALSE |
| 2d - stereo | Sticks | -0.023 | 0.03 | 0.499 | FALSE |

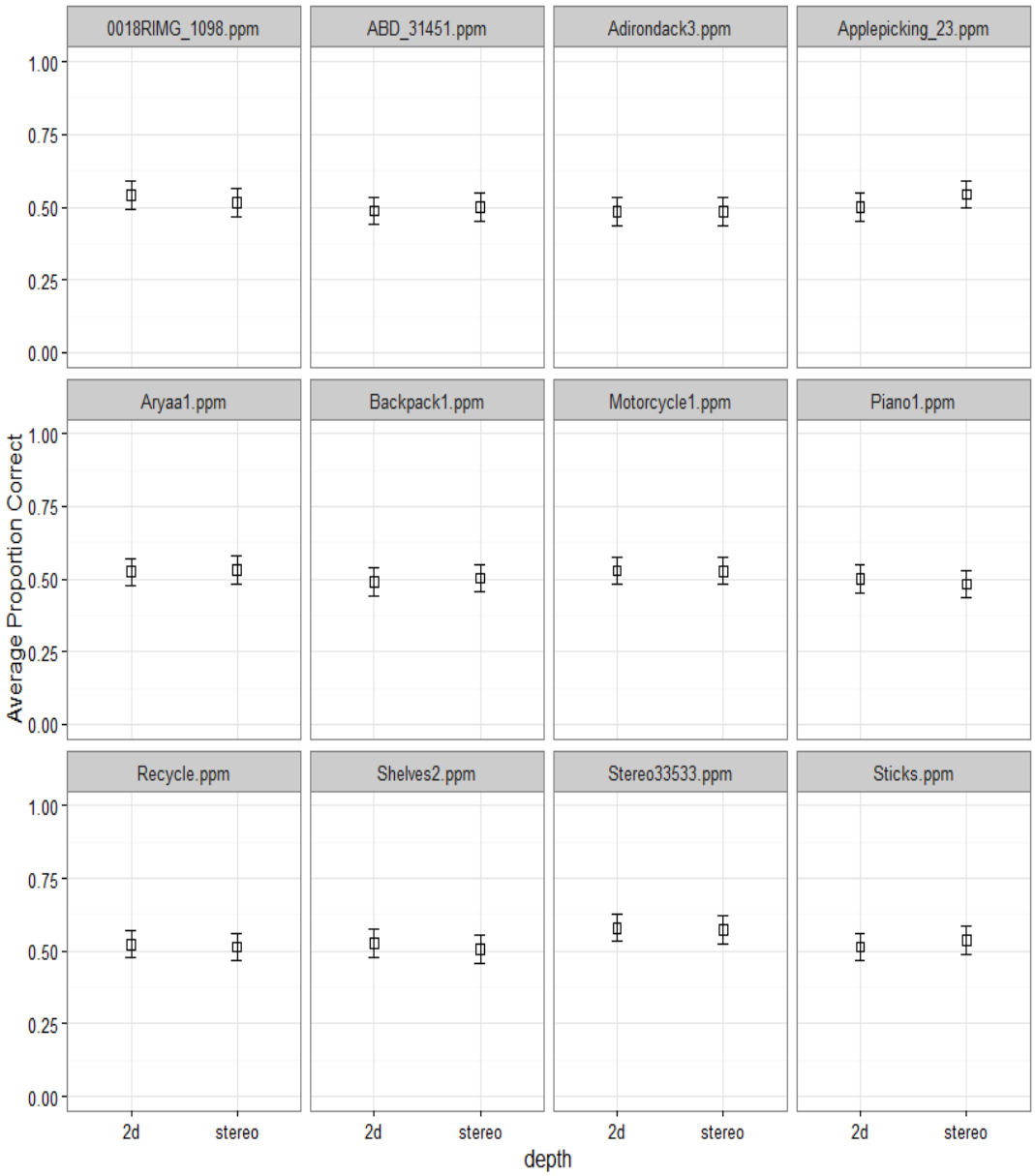



**Figure 5.7:** Average proportion correct for 2D and stereo condition when VDCM 1.0.7 (collapsed across 6bpp and 4bpp) compression was used. Error bars show ±1 SE.

**Table 7:** Comparison between 2D and Stereo for each image with VDC-M 1.0.7 (6:1) compression.

| | Experiment2_hypothesis 2.2_2DvsStereo_4bpp_two tail | | | | |
|---|---|---|---|---|---|
| contrast | ref | estimate | SE | p.value | sig |
| 2d – stereo | 0018RIMG_1098 | 0.023 | 0.05 | 0.632 | FALSE |
| 2d – stereo | ABD_31451 | -0.014 | 0.05 | 0.774 | FALSE |
| 2d – stereo | Adirondack3 | -0.005 | 0.05 | 0.924 | FALSE |
| 2d – stereo | Applepicking_23 | -0.073 | 0.05 | 0.126 | FALSE |
| 2d – stereo | Aryaa1 | -0.027 | 0.05 | 0.567 | FALSE |
| 2d – stereo | Backpack1 | -0.036 | 0.05 | 0.445 | FALSE |
| 2d – stereo | Motorcycle1 | 0.023 | 0.05 | 0.633 | FALSE |
| 2d – stereo | Piano1 | 0.068 | 0.05 | 0.151 | FALSE |
| 2d – stereo | Recycle | -0.036 | 0.05 | 0.445 | FALSE |
| 2d – stereo | Shelves2 | 0.086 | 0.05 | 0.069 | FALSE |
| 2d – stereo | Stereo33533 | 0.036 | 0.05 | 0.427 | FALSE |
| 2d - stereo | Sticks | -0.027 | 0.05 | 0.566 | FALSE |

Inspection of Figure 5.7 shows that performance was typically near chance levels for all images. We expect that detection rate for artefacts should be higher for compression at 4 bpp compared to 6 bpp. Thus, Hypothesis 2.2 proposed that there would be differences between stereo and 2D conditions for 4 bpp (6:1) compression. The results of these comparisons are shown in Table 7. Similar to the results in the full dataset there were no significant differences between 2D vs stereo conditions for any image. Figure 5.8 plots proportion correct data for 4 bpp

compression for each image in both 2D and stereo conditions. Note that for all images' performance is near 0.5, so near chance performance, this floor effect limits the ability to test for difference between conditions even in the 4 bpp case.

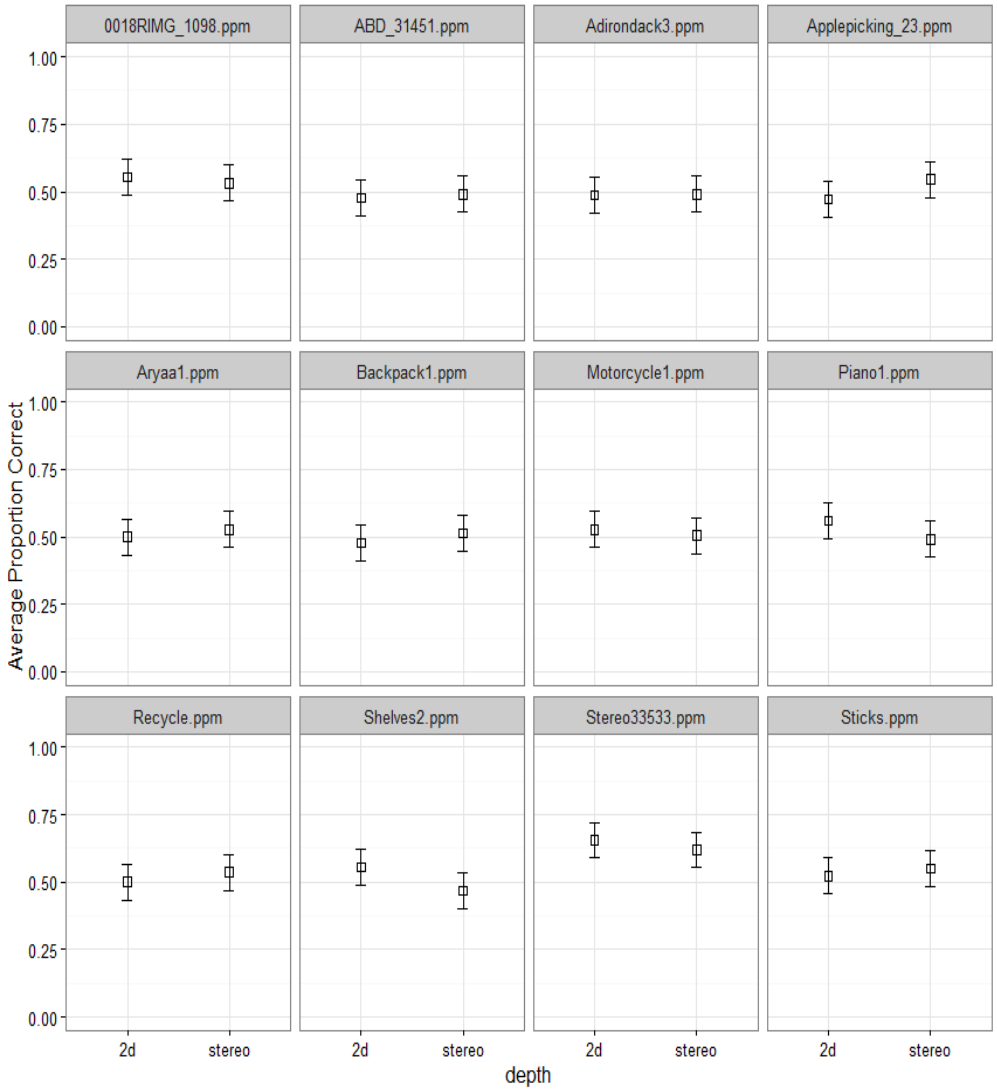



**Figure 5.8:** Average proportion correct for 2D versus stereo condition when VDCM 1.0.7 (4bpp) compression is used. Error bars shows ±1 SE.

**Table 8:** Comparison between 4 bpp (6:1) versus 6 bpp (4:1) compression for each image in stereo condition.

| | Experiment2_hypothesis 2.3_4bppVS6bpp_stereo_right tail | | | | |
|---|---|---|---|---|---|
| contrast | ref | estimate | SE | p.value | sig |
| 4bpp – 6bpp | 0018RIMG_1098 | 0.032 | 0.05 | 0.252 | FALSE |
| 4bpp – 6bpp | ABD_31451 | -0.018 | 0.05 | 0.649 | FALSE |
| 4bpp – 6bpp | Adirondack3 | 0.014 | 0.05 | 0.387 | FALSE |
| 4bpp – 6bpp | Applepicking_23 | 0.005 | 0.05 | 0.462 | FALSE |
| 4bpp – 6bpp | Aryaa1 | -0.009 | 0.05 | 0.576 | FALSE |
| 4bpp – 6bpp | Backpack1 | 0.022 | 0.05 | 0.317 | FALSE |
| 4bpp – 6bpp | Motorcycle1 | -0.045 | 0.05 | 0.830 | FALSE |
| 4bpp – 6bpp | Piano1 | 0.018 | 0.05 | 0.351 | FALSE |
| 4bpp – 6bpp | Recycle | 0.045 | 0.05 | 0.170 | FALSE |
| 4bpp – 6bpp | Shelves2 | -0.077 | 0.05 | 0.948 | FALSE |
| 4bpp – 6bpp | **Stereo33533** | 0.091 | 0.05 | **0.026** | **TRUE** |
| 4bpp – 6bpp | Sticks | 0.027 | 0.05 | 0.283 | FALSE |

The effects of VDC-M 1.0.7 compression level were explicitly addressed in Hypothesis 2.3 for the subset of stereo conditions and Hypothesis 2.4 for the 2D conditions. As I expected that the performance for 4 bpp (6:1) compression would be worse than 6 bpp (4:1) compression, a one tail pairwise comparison was conducted to test the hypotheses. Table 8 shows the results of this comparison for the stereo conditions (Hypothesis 2.3) and the corresponding data is plotted in Figure 5.9. The only image where the performance at 6bpp was significantly

different from 4bpp was Stereo33533. From the plot we can see that the difference between the mean 4bpp and 6bpp for Stereo33533 was comparatively larger than for the other images.

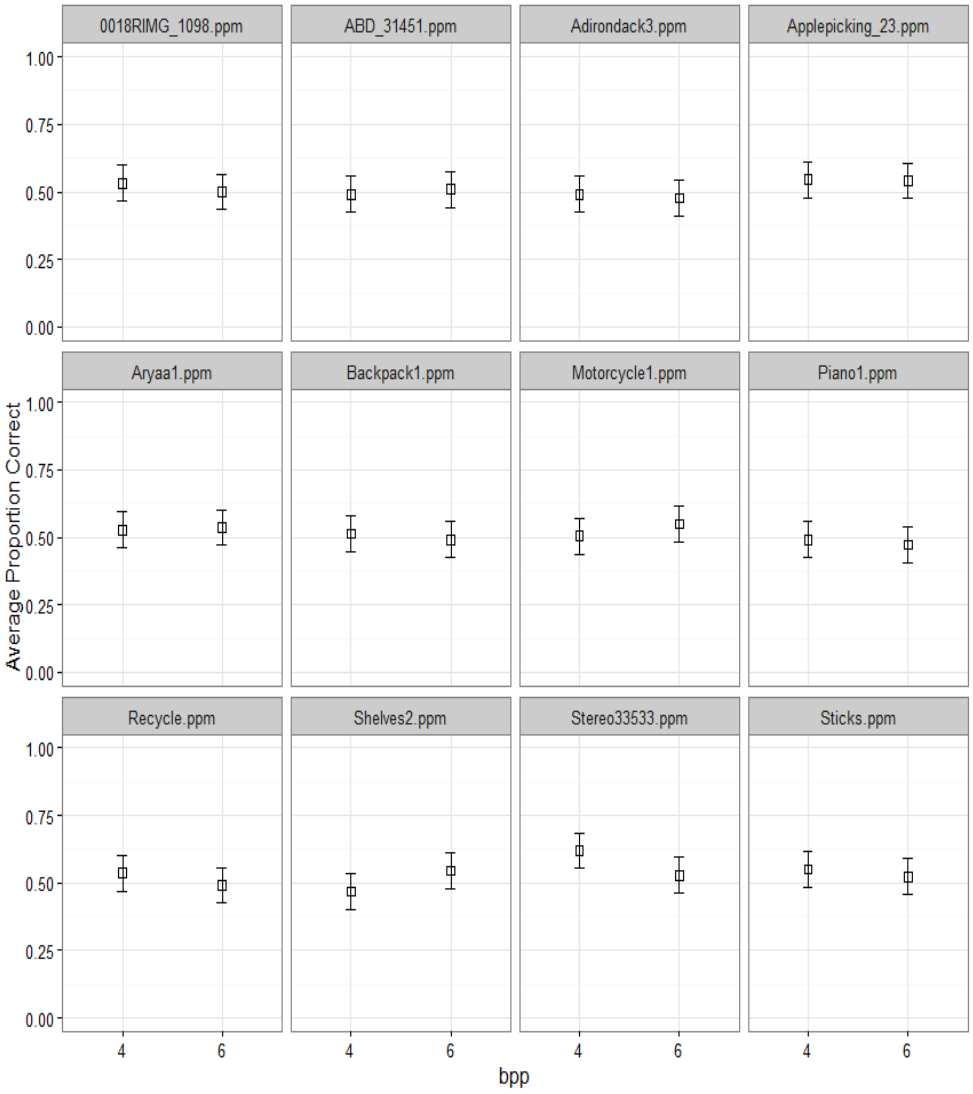



**Figure 5.9:** Average proportion correct for 4bpp versus 6bpp in stereo condition when VDC-M 1.0.7 compression was used. Error bars shows ±1 SE.

**Table 9:** Comparison between 4 bpp (6:1) versus 6bpp (4:1) compression for each image in 2D condition.

| | Experiment2_hypothesis 2.4_4bppVS6bpp_2D_right tail | | | | |
|---|---|---|---|---|---|
| contrast | ref | estimate | SE | p.value | sig. |
| 4 bpp – 6 bpp | 0018RIMG_1098 | 0.027 | 0.05 | 0.283 | FALSE |
| 4 bpp – 6 bpp | ABD_31451 | -0.018 | 0.05 | 0.649 | FALSE |
| 4 bpp – 6 bpp | Adirondack3 | 0.005 | 0.05 | 0.462 | FALSE |
| 4 bpp – 6 bpp | Applepicking_23 | -0.055 | 0.05 | 0.874 | FALSE |
| 4 bpp – 6 bpp | Aryaa1 | -0.050 | 0.05 | 0.853 | FALSE |
| 4 bpp – 6 bpp | Backpack1 | -0.027 | 0.05 | 0.717 | FALSE |
| 4 bpp – 6 bpp | Motorcycle1 | -0.005 | 0.05 | 0.538 | FALSE |
| 4 bpp – 6 bpp | **Piano1** | 0.118 | 0.05 | **0.006** | **TRUE** |
| 4 bpp – 6 bpp | Recycle | -0.046 | 0.05 | 0.830 | FALSE |
| 4 bpp – 6 bpp | Shelves2 | 0.055 | 0.05 | 0.126 | FALSE |
| 4 bpp – 6 bpp | **Stereo33533** | 0.150 | 0.05 | **0.0006** | **TRUE** |
| 4 bpp – 6 bpp | Sticks | 0.018 | 0.05 | 0.351 | FALSE |

I did a similar comparison of the performance of VDCM 1.0.7 for 4bpp vs 6bpp in the 2D conditions to test hypothesis 2.4. The results of these comparisons are shown in Table 9 and the difference between performance can be visualized in Figure 5.10 by comparing the mean for 4bpp vs 6bpp. For two images, Piano1 and Stereo33533, the compression algorithm performed significantly worse (artefacts were more visible) at 4 bpp than at 6 bpp**.**

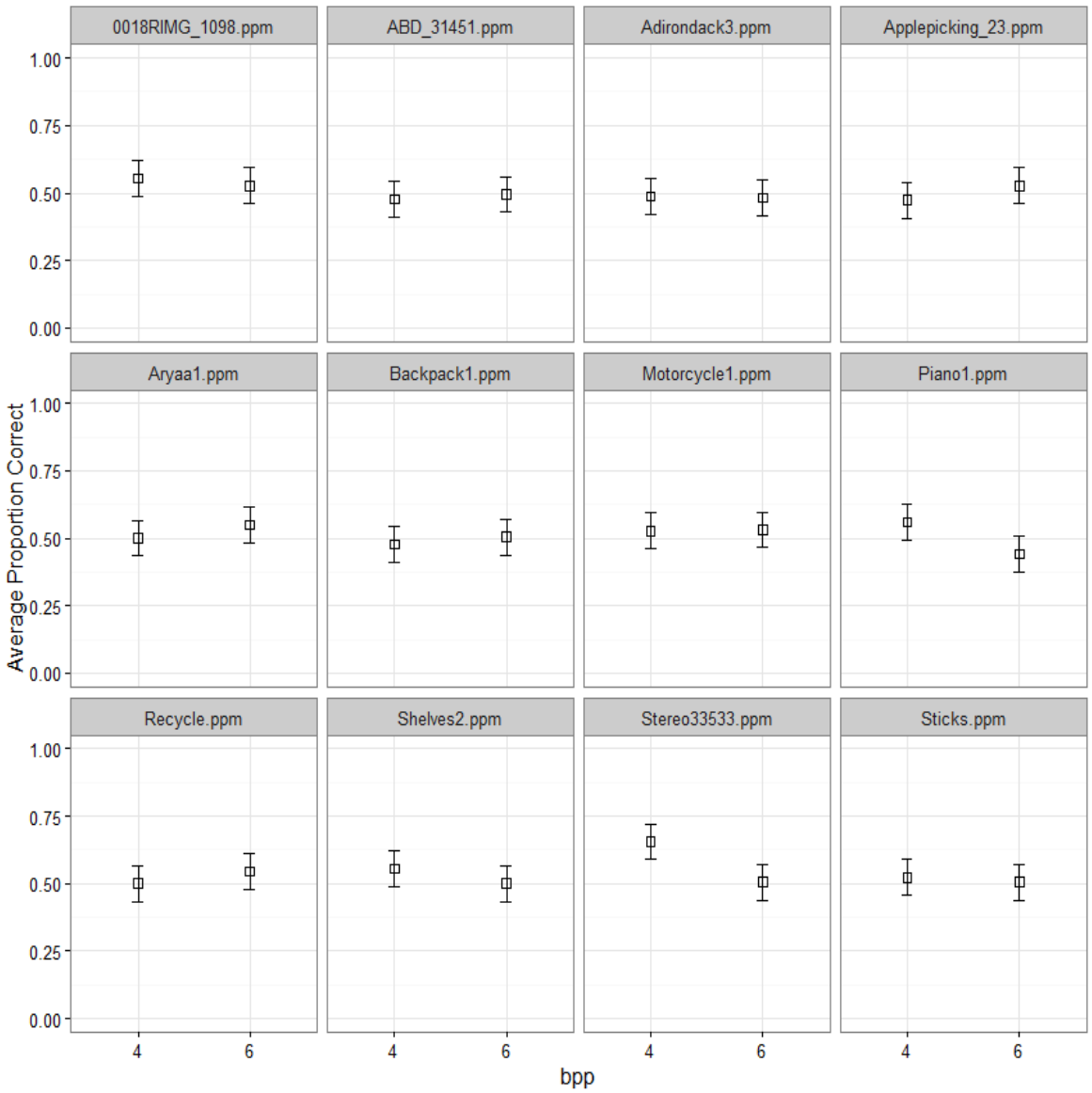



**Figure 5.10:** Average proportion correct for 4bpp and 6bpp in 2D condition when VDCM 1.0.7 compression was used. Error bars shows ±1 SE.

Only two images (Figure 5.11) showed significant effects of compression: Stereo33533 for both stereo and 2D cases and Piano1 for 2D images. The distinguishing feature of Stereo33533 is that it is a natural scene which contains a fine green leafy pattern and in Piano1 the floor carpet has a noise like texture. Due to these features, the codec faced more difficulty in compressing these images compared to others which made it possible to see the difference between 4 bpp and 6 bpp for these two images. Most of the remaining images were relatively easy for the codec to compress and performance was near chance at both 4 bpp and 6 bpp.

**Figure 5.11:** Piano1 and Stereo33533 (Left eye)

# Chapter 6
# Discussion, Conclusion and Future Work

## 6.1 Discussion

During my thesis I tested the perception of compression artefacts in stereoscopic display for two different sets of images in both 2D and 3D (stereo) conditions. For the 1st experiment image set, I found that for three images (FemaleHorseFly, Noise and Tools) in all conditions DSC 1.2 faced difficulties performing as a visually lossless codec and for 'Tools' this was consistent with VESA 2D testing of these images (Allison et al. 2017).

The main objectives were to determine the effects of stereoscopic depth and symmetry of artefacts between half images. I found that for two specific images (Mandrill and Peacock) there was a significant effect of interaction between the depth (2D or stereo) and compression (same or different) on the visibility of compression artefact. For 4 images (CircularPattern26, Landscape102, Mandrill and Tools) DSC 1.2 performed significantly better for stereo images than 2D (i.e., stereo vision was silencing the artefact perception for these images) but for Clipboard artefacts were significantly more visible in stereo images than 2D. In 5 images (FemaleHorseFly, Landscape 102, MosaicBroadcom, Noise, Tools) artefacts were significantly less perceptible in images with different compression compared to images with same compression. The exception again was Clipboard

where the codec performed significantly better for same compression than for different compression. Thus, I conclude that both disparity and symmetry of compression artefacts affect artefact visibility which is consistent with the findings in (Chen, Bovik, and Cormack 2011).

For the 2nd set of images VDCM 1.0.7 performed as visually lossless for all images and in most of the cases the mean was around 50% which indicates guessing of the observers in detecting flicker images that is, there was a floor effect. The floor effect in Experiment 2 does not call into question the practical relevance of the results in Experiment 1. Experiment 1 is aimed at determining, for the (relatively rare) cases where artefacts are visible, whether stereoscopic viewing or compression symmetry are factors in the detectability of these artefacts. The fact that the codec normally works well does not imply that it may or may not be sensitive to disparity and compression differences.

In the 2nd Experiment, the performance of the codec was not significantly different between the 2D and stereo images at both levels (4 bpp, and 6 bpp) of compression. I compared the artefact perception between two compression levels for stereo and 2D images which showed that in both 2D and stereo, 6 bpp performs significantly better than 4 bpp for “Stereo33533” and in 2D, 6 bpp performs significantly better than 4 bpp for “Piano 1”. As most of the images in the 2nd image set exhibited flicker detection rates was near chance and performance for 2D and stereo images did not differ significantly, the images used for 2nd experiment were

comparatively harder for the subjects to detect the flickering but easier for the codec to compress. This can be seen in the PSNR (Appendix A) which are relatively high. So, I cannot come to a conclusion for the 2nd set of images like the 1st set of images for which I found significant difference between 2D and 3D performance. As in many applications we use different types of images and some types of images could be easier for the codec and some could be harder for the codec to compress, it is important to choose the representative images (comparatively harder for the codec and different types) to test a codec like VDC-M to get a generalized conclusion. To get a conclusive answer, an experiment with images which are challenging for the codec is necessary. One of the ways of getting challenging images is rendering images. For this, some difficult images can be selected from the 1st set and we can get a crop from each of selected images and can render stereo images using those crops as textures. The compression quality can be controlled by varying the slice size of VDC-M, normally a larger slice size provides better quality images (less artefacts). But as we are interested to see the viewer response on the images with more artefacts, we can compress images more harshly by varying (reducing) the slice size. A subjective experiment with the rendered images along with some of the 2nd set images (which are comparatively harder for the codec, e.g., Stereo 33533) may help us to find the differences between stereoscopic and 2D presentation. It will be necessary to test the codec

with more challenging, engineered and rendered images to come to a clear conclusion.

## 6.2 Conclusions

In my thesis I tried to assess the human perception of image compression artefacts. It is always desirable that the visually lossless property of a compression codec be obtained because we know in many cases compression artefacts may lead to misunderstanding of information in images e.g. in health and treatment related images, and in images related to national security (defense). Moreover, 3D display is becoming more popular, and people are getting more interested in VR (Virtual Reality) using HMDs (Head Mounted Displays). In all these cases, we would prefer to use a compression codec which can provide visually lossless images. As objective measurement is not sufficient to assess the visually lossless property, it is necessary to assess any compression codec subjectively to verify the visually lossless property.

In my thesis I did subjective assessment for DSC 1.2 and VDC-M 1.0.7 with two sets of images to assess their performance in 2D vs 3D and found that for one set of images (2D with disparity) DSC performs better in 3D compared to 2D and for another set of images (the stereoscopic images), no conclusive result was obtained and further investigation is needed to come to a conclusion.

## 6.3 Future Work

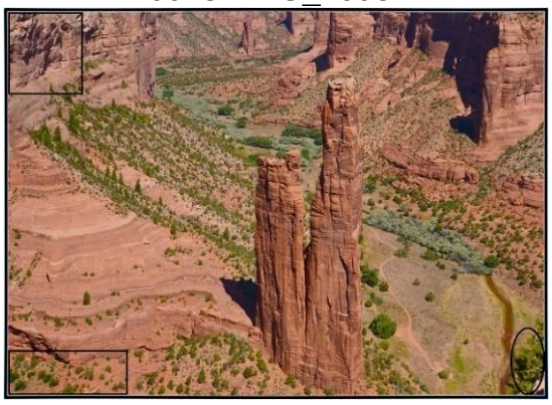
For virtual reality applications, Head Mounted Displays are used. The optics used in HMDs can cause Chromatic Aberration. When light passes through a lens, the lens can fail to converge different wavelengths of light at the same point due to variation in refractive indices with wavelength. This is called chromatic aberration. Due to chromatic aberration, colored fringes become visible around the object and are more apparent in edges of the image compared to the center. In most modern HMDs, the image is pre-transformed (distorted) so that the lens can provide a corrected image after chromatic aberration. Using this distortion correction, we can reduce chromatic aberration but cannot remove it. Because each color channel comprises of a range of visible wavelengths each of which is refracted by a different amount by the lens, we can distort the images for each color channel (e.g., R, G, B) to bring the peak frequencies back into spatial alignment but it may not possible to compensate for the aberration within each color channel.

The flicker paradigm can be used to do the assessment of artefacts in the compressed image with chromatic aberration. In order to see the artefacts in chromatic aberration images, I distorted the 2$^{nd}$ experiment image set with barrel distortion by varying the distortion coefficient for each color channel, as a result I got some images with chromatic aberration and then I compressed those images with VDCM 1.0.7 6:1 compression and then inverse distorted the compressed images.

I compared those compressed images with the uncompressed images applying the flicker protocol; in some images (e.g., 0018RIMG_1098, ABD_31451, Applepicking_23, Stereo33533) I saw that flicker was more apparent for the compressed image with chromatic aberration than the compressed image without chromatic aberration. Figure 6.1 shows those images in full size and Figure 6.2 shows the cropped portions, the regions highlighted with a black circle or black rectangular show more flicker for compressed images with chromatic aberration compared to compressed images without chromatic aberration. But I cannot make firm conclusions regarding this from this pilot testing, to get a concrete decision subjective assessment is necessary.

0018RIMG_1098

Stereo33533



ABD_31451

Applepicking_23

**Figure 6.1:** Full images which show more flicker in the compressed images with chromatic aberration compared to compressed images without chromatic aberration

0018RIMG_1098 Stereo33533

ABD_31451

Applepicking_23

**Figure 6.2:** Cropped region of the images which show more flicker in the compressed images with chromatic aberration compared to compressed images without chromatic aberration.

# Bibliography


Agresti, Alan. 2007. “Random Effects: Generalized Linear Mixed Models.” In *An Introduction to Categorical Data Analysis*.

Allison, Robert S., and Laurie M. Wilcox. 2015. “Perceptual Tolerance to Stereoscopic 3D Image Distortion.” *ACM Trans. Appl. Percept.* 12 (3): 10:1–10:20. https://doi.org/10.1145/2770875.

Allison, Robert S., Laurie M. Wilcox, Wei Wang, David M. Hoffman, Yuqian Hou, James Goel, Lesley Deas, and Dale Stolitzka. 2017. “75-2: Invited Paper: Large Scale Subjective Evaluation of Display Stream Compression.” *SID Symposium Digest of Technical Papers* 48 (1): 1101–4. https://doi.org/10.1002/sdtp.11838.

Anderson, Patricia A., and J. Anthony Movshon. 1989. “Binocular Combination of Contrast Signals.” *Vision Research* 29 (9): 1115–32. https://doi.org/10.1016/0042-6989(89)90060-6.

Barth, N.R. 2010. “Ringing Artifact Example.” https://en.wikipedia.org/wiki/Ringing_artifacts#/media/File:Ringing_artifact_example.png.

Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1): 289–300.

Blake, R., and E. Levinson. 1977. “Spatial Properties of Binocular Neurones in the Human Visual System.” *Experimental Brain Research* 27 (2): 221–32. https://doi.org/10.1007/BF00237700.

Brunnström, K., R. S. Allison, D. M. Chandler, H. Colett, P. Corriveau, S. Daly, J. Goel, et al. 2017. “Industry and Business Perspectives on the Distinctions between Visually Lossless and Lossy Video Quality: Mobile and Large Format Displays.” *Electronic Imaging* 2017 (14): 118–33. https://doi.org/10.2352/ISSN.2470-1173.2017.14.HVEI-131.

Campbell, F. W., and D. G. Green. 1965. “Monocular versus Binocular Visual Acuity.” *Nature* 208 (5006): 191–92. https://doi.org/10.1038/208191a0.

Campisi, P., P. Le Callet, and E. Marini. 2007. “Stereoscopic Images Quality Assessment.” In *2007 15th European Signal Processing Conference*, 2110–14.

Chen, M. J., A. C. Bovik, and L. K. Cormack. 2011. “Study on Distortion Conspicuity in Stereoscopically Viewed 3D Images.” In *2011 IEEE 10th IVMSP Workshop: Perception and Visual Signal Analysis*, 24–29. https://doi.org/10.1109/IVMSPW.2011.5970349.

Choi, Lark Kwon, Lawrence K. Cormack, and Alan C. Bovik. 2015. “Motion Silencing of Flicker Distortions on Naturalistic Videos.” *Signal Processing: Image Communication*, Recent Advances in Vision Modeling for Image and Video Processing, 39 (November): 328–41. https://doi.org/10.1016/j.image.2015.03.006.

Descampe, Antonin, Joachim Keinert, Thomas Richter, Siegfried Fößel, and Gaël Rouvroy. 2017. “JPEG XS, a New Standard for Visually Lossless Low-Latency Lightweight Image Compression.” In *Applications of Digital Image Processing XL*, 10396:103960M. International Society for Optics and Photonics. https://doi.org/10.1117/12.2273625.

Dhawan, Sachin. 2011. “A Review of Image Compression and Comparison of Its Algorithms.” *International Journal of Electronics & Communication Technology, IJECT* 2 (1): 22–6.

Flickr website. 2018. “Flickr Images.” https://www.flickr.com/photos/tags/flicker/.

Hoffman, David M., Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. 2008. “Vergence–Accommodation Conflicts Hinder Visual Performance and Cause Visual Fatigue.” *Journal of Vision* 8 (3): 33–33. https://doi.org/10.1167/8.3.33.

Hoffman, David M., and Dale Stolitzka. 2014. “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.

Howard, Ian P., and Brian J. Rogers. 2012. “Binocular Summation, Masking, and Transfer.” In *Perceiving in Depth. Volume 2, Stereoscopic Vision*, 107–47. Oxford : Oxford University Press.

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

International Telecommunications Union. (2011). *BT 1886*, “Reference Electro-Optical Transfer Function for Flat Panel Displays Used in HDTV Studio Production.” https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.1886-0-201103-I!!PDF-E.pdf.

International Telecommunications Union. (2018). *BT 2100*, “Image Parameter Values for High Dynamic Range Television for Use in Production and International Programme Exchange.” https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.2100-2-201807-I!!PDF-E.pdf.

Jacobson, Natan, Vijayaraghavan Thirumalai, Rajan Joshi, and James Goel. 2017. “A New Display Stream Compression Standard under Development in VESA.” In *Applications of Digital Image Processing XL*, 10396:103960U. International Society for Optics and Photonics. https://doi.org/10.1117/12.2274595.

Lambooij, Marc, Marten Fortuin, Ingrid Heynderickx, and Wijnand IJsselsteijn. 2009. “Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review.” *Journal of Imaging Science and Technology* 53 (3): 30201–1.

Legge, Gordon E. 1984. “Binocular Contrast Summation—I. Detection and Discrimination.” *Vision Research* 24 (4): 373–83. https://doi.org/10.1016/0042-6989(84)90063-4.

Marcos, Susana, Esther Moreno, and Rafael Navarro. 1999. “The Depth-of-Field of the Human Eye from Objective and Subjective Measurements.” *Vision Research* 39 (12): 2039–49. https://doi.org/10.1016/S0042-6989(98)00317-4.

Maddox, E. E. 1886. “Investigations in the Relation between Convergence and Accommodation of the Eyes.” *Journal of Anatomy and Physiology* 20 (Pt 3): 475–508.

McIntire, John P., Paul R. Havig, and Eric E. Geiselman. 2014. “Stereoscopic 3D Displays and Human Performance: A Comprehensive Review.” *Displays* 35 (1): 18–26. https://doi.org/10.1016/j.displa.2013.10.004.

McKee, Suzanne P., Stanley A. Klein, and Davida Y. Teller. 1985. “Statistical Properties of Forced-Choice Psychometric Functions: Implications of Probit Analysis.” *Perception & Psychophysics* 37 (4): 286–98. https://doi.org/10.3758/BF03211350.

Poirson, A. B., and B. A. Wandell. 1997. “Pattern-Color Separable Pathways Predict Sensitivity to Simple Colored Patterns.” *Ophthalmic Literature* 1 (50): 50.

Posterization Artifacts, 2018. https://svi.nl/PosterizationArtifacts.

Punchihewa, Amal, and Donald G. Bailey. 2002. “Artefacts in Image and Video Systems; Classification and Mitigation.” In *Proceedings of Image and Vision Computing New Zealand*, 197–202.

Regan, David, and Rob Gray. 2009. “Binocular Processing of Motion: Some Unresolved Questions.” *Spatial Vision* 22 (1): 1–43. https://doi.org/10.1163/156856809786618501.

SAS Work Shop. 2018. “LSMEANS.” http://webpages.uidaho.edu/cals-statprog/sas/workshops/glm/handout2.1_glm.pdf.

Scharstein, Daniel, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nešić, Xi Wang, and Porter Westling. 2014. “High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth.” In *Pattern Recognition*, edited by Xiaoyi Jiang, Joachim Hornegger, and Reinhard Koch, 31–42. Lecture Notes in Computer Science. Springer International Publishing.

Seuntiens, Pieter, Lydia Meesters, and Wijnand Ijsselsteijn. 2006. “Perceived Quality of Compressed Stereoscopic Images: Effects of Symmetric and Asymmetric JPEG Coding and Camera Separation.” *ACM Transactions on Applied Perception (TAP)* 3 (2): 95–109.

Society of Motion Picture and Television Engineers (SMPTE) (2014), “ST 2084:2014 - SMPTE Standard - High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays.” *ST 2084:2014*, August, 1–14. https://doi.org/10.5594/SMPTE.ST2084.2014.

Stolitzka, Dale. 2017. “Personal Communication.”

Tam, Wa James, Filippo Speranza, Sumio Yano, Koichi Shimono, and Hiroshi Ono. 2011. “Stereoscopic 3D-TV: Visual Comfort.” *IEEE Transactions on Broadcasting* 57 (2): 335–346.

Video Electronic Standard Association (2018). “VESA Display Compression-M (VDC-M) Standard.” www.vesa.org.

Walls, and A. S. MacInnis. 2016. “VESA Display Stream Compression for Television and Cinema Applications.” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 6 (4): 460–70. https://doi.org/10.1109/JETCAS.2016.2602009.

Walls, and A.S. MacInnis. 2014. “27.4L: Late-News Paper: VESA Display Stream Compression: An Overview.” *SID Symposium Digest of Technical Papers* 45 (1): 360–63. https://doi.org/10.1002/j.2168-0159.2014.tb00097.x.

Wang, Xu, Mei Yu, You Yang, and Gangyi Jiang. 2009. “Research on Subjective Stereoscopic Image Quality Assessment.” In *Multimedia Content Access: Algorithms and Systems III*, 7255:725509. International Society for Optics and Photonics.

Wichmann, Felix A., and N. Jeremy Hill. 2001. “The Psychometric Function: I. Fitting, Sampling, and Goodness of Fit.” *Perception & Psychophysics* 63 (8): 1293–1313. https://doi.org/10.3758/BF03194544.

Wilkinson, G. N., and C. E. Rogers. 1973. “Symbolic Description of Factorial Models for Analysis of Variance.” *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 22 (3): 392–99. https://doi.org/10.2307/2346786.

Yang, Jiachen, Chunping Hou, Yuan Zhou, Zhuoyun Zhang, and Jichang Guo. 2009. “Objective Quality Assessment Method of Stereo Images.” In *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2009*, 1–4. IEEE.

Yang, Shunnan, and James E. Sheedy. 2011. “Effects of Vergence and Accommodative Responses on Viewer’s Comfort in Viewing 3D Stimuli.” In *Stereoscopic Displays and Applications XXII*, 7863:78630Q. International Society for Optics and Photonics. https://doi.org/10.1117/12.872546.

Zaroff, Charles M., Magosha Knutelska, and Thomas E. Frumkes. 2003. “Variation in Stereoacuity: Normative Description, Fixation Disparity, and the Roles of Aging and Gender.” *Investigative Ophthalmology & Visual Science* 44 (2): 891–900. https://doi.org/10.1167/iovs.02-0361.

Zhang, X., and B. A. Wandell. 1997. “A Spatial Extension of CIELAB for Digital Color-Image Reproduction.” *Journal of the Society for Information Display* 5 (1): 61–63. https://doi.org/10.1889/1.1985127.

# Appendix

**Table 10:** PSNR of group 1 images DSC 1.2 compression

| Image | PSNR value |
|---|---|
| Barbara | 43.09027 |
| CircularPattern26 | 34.83199 |
| Clipboard | 43.46617 |
| FemaleHorseFly | 42.28563 |
| HintergroundMusik | 34.16949 |
| Landscape102 | 31.91763 |
| Mandrill | 34.25364 |
| MosaicBroadcom | 34.96823 |
| MysticMountain | 40.179 |
| Noise | 27.40965 |
| Peacock | 37.3827 |
| Tools | 36.16092 |

**Table 11:** PSNR of group 2 images VDCM 1.0.7 compression

| | VDCM 1.0.7 4:1 compression | | VDCM 1.0.7 6:1 compression | |
|---|---|---|---|---|
| Image | Left | Right | Left | Right |
| Applepicking_23 | 47.7089 | 48.42846 | 41.98305 | 42.97548 |
| ABD_31451 | 48.12504 | 48.26707 | 42.87653 | 43.11858 |
| Sticks | 50.9261 | 49.30327 | 48.40558 | 46.01847 |
| Recycle | 53.01836 | 53.02093 | 50.21592 | 50.22285 |
| Piano1 | 50.82574 | 50.70926 | 47.59869 | 47.37512 |
| Shelves2 | 51.76947 | 51.88736 | 49.01531 | 49.07846 |
| Aryaa1 | 48.39568 | 48.00381 | 44.65459 | 43.87826 |
| Backpack1 | 50.55872 | 50.313 | 47.96689 | 47.56243 |
| Stereo33533 | 43.54409 | 43.24134 | 36.7356 | 36.42986 |
| Adirondack3 | 51.87927 | 51.9694 | 49.34984 | 49.43008 |
| Motorcycle1 | 50.78851 | 50.81398 | 48.26455 | 48.28291 |
| 0018RIMG_1098 | 49.20521 | 51.68001 | 41.75922 | 46.61259 |