

# Evaluating 360° media experiences

*Andrew MacQuarrie*

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
**Doctor of Engineering**

Department of Computer Science  
University College London  
December 13, 2018

I, Andrew MacQuarrie, confirm that the work presented herein is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

# Abstract

360° media experiences have existed for centuries. Viewing painted panoramas, such as those displayed in the 18th-century rotunda in Leicester Square, was a popular Georgian pastime. Recent advances in capture, processing and display technology have created a surge of interest in the medium, with millions of people now viewing captured 360° media immersively. Despite the popularity of 360° media experiences, there are still substantial technical issues associated with production and distribution, and little research has been done that explores the end-user experience. As these experiences become commonplace, understanding the impact of such media becomes critical.

In this work, two user studies were conducted that investigated the effects of 360° media of different forms. The first study looked at the impact of the display type when viewing cinematic virtual reality captured as 360° video. The study used three display types: a head-mounted display (HMD); a standard 16:9 TV; and a focus-plus-context display. Several metrics were explored, including spatial awareness, memory and narrative engagement. The second study investigated the impact of different transition types when exploring static scenes captured as multi-view 360° images in a HMD. The three transitions investigated were a linear movement through a 3D model of the scene, an instantaneous teleportation, and an image-based warp using Möbius transformations. Metrics investigated included spatial awareness, preference, and several subjective qualities such as the feeling of moving through the space.

Additionally, an enabling technology for such experiences was investigated. Object removal in 360° images was explored in detail, with extensions for video described for simple cases. Taken together, these three projects further our current understanding of how 360° media can be implemented, and examine some of the most critical aspects of how users engage with these experiences.

# Acknowledgements

This work was supported in part by grant EP/G037159/1 for University College London's Virtual Environments, Imaging & Visualisation Doctoral Training Centre from the UK Engineering and Physical Sciences Research Council (EPSRC) and BBC R&D.

I would like to thank my primary supervisor, Anthony Steed. His intelligent guidance made this project possible, while his approach made it fun. The VR group at UCL has also contributed greatly to my learning and enjoyment throughout, with special thanks to David Swapp, Sebastian Friston, Ye Pan, Maria Murcia Lopez, David Walton, Jacob Thorn and Ben Congdon. Thanks also to Jozef Dobos and Will Steptoe for all their support when I first joined the team, as well as Dave Twistleton for all his help with physical setups.

Thanks also to BBC R&D, and in particular my industrial supervisor Graham Thomas, as well as Zillah Watson and Alia Sheikh, for all their help throughout, and for allowing me to get involved in various filming projects. Thanks also to BBC R&D, Peter Boyd Maclean and Matterport, for all their guidance, and allowing me to use their content.

Finally, thanks to my father, brother and Simon, without whom this work would not have been possible.

# Contents

<b>1</b>	<b>Introduction</b>	<b>12</b>
1.1	Research problem . . . . .	13
1.2	Scope . . . . .	14
1.3	Contributions . . . . .	15
1.4	Publications . . . . .	15
1.5	Structure . . . . .	16
<b>2</b>	<b>Background</b>	<b>18</b>
2.1	Panoramic media . . . . .	19
2.1.1	Panoramic media capture . . . . .	19
2.1.2	Cinematic virtual reality . . . . .	19
2.1.3	Multi-view panoramic media . . . . .	20
2.1.4	Panoramic media editing . . . . .	21
2.2	Immersive displays . . . . .	26
2.2.1	Head-mounted displays . . . . .	26
2.2.2	Projection-based displays . . . . .	29
2.3	Evaluation . . . . .	35
2.3.1	Evaluation of cinematic virtual reality experiences . . . . .	35
2.3.2	Evaluation of multi-view 360° experiences . . . . .	39
2.3.3	Simulator sickness . . . . .	42
<b>I</b>	<b>Media production and acquisition</b>	<b>45</b>
<b>3</b>	<b>Media acquisition</b>	<b>47</b>
3.1	BBC in collaboration with UCL . . . . .	47
3.1.1	Test documentary . . . . .	47

3.1.2	The Resistance of Honey . . . . .	50
3.2	BBC and external companies . . . . .	51
3.3	YouTube and the growth of 360° video . . . . .	53
3.4	Multi-view 360° media . . . . .	53
3.5	Conclusion . . . . .	54
<b>4</b>	<b>Object removal in panoramic media</b>	<b>56</b>
4.1	FOV expansion and graph cuts . . . . .	56
4.1.1	Algorithm formulation . . . . .	56
4.1.2	Results . . . . .	57
4.1.3	Limitations . . . . .	58
4.2	Extensions & refinements . . . . .	62
4.2.1	Retargeting techniques . . . . .	62
4.2.2	Tripod removal . . . . .	64
4.3	Inpainting . . . . .	65
4.3.1	Inpainting in equirectangular . . . . .	65
4.3.2	Straight line preserving projections . . . . .	68
4.3.3	Limitations . . . . .	70
4.4	Video . . . . .	71
4.5	Conclusion . . . . .	73
<b>II</b>	<b>360° media evaluation</b>	<b>75</b>
<b>5</b>	<b>User study: evaluating the effect of display type on the viewing experience for panoramic video</b>	<b>77</b>
5.1	Study design . . . . .	78
5.1.1	Subjects . . . . .	79
5.1.2	Experimental conditions . . . . .	79
5.1.3	Stimuli . . . . .	81
5.1.4	Hypotheses . . . . .	83
5.1.5	Measures . . . . .	83
5.1.6	Procedure . . . . .	86

5.2	Results and discussion . . . . .	88
5.3	Limitations . . . . .	99
5.4	Conclusion . . . . .	101
<b>6</b>	<b>User study: The effect of transition type in multi-view 360° media</b>	<b>104</b>
6.1	Experimental Design . . . . .	105
6.1.1	Stimuli . . . . .	106
6.1.2	Transition Types . . . . .	107
6.1.3	Hypotheses . . . . .	110
6.1.4	Experimental Setup . . . . .	113
6.1.5	Participants . . . . .	115
6.1.6	Experimental Procedure . . . . .	115
6.2	Results . . . . .	119
6.2.1	Spatial Awareness . . . . .	119
6.2.2	Subjective Measures . . . . .	123
6.2.3	Preference . . . . .	125
6.2.4	Movement Profile . . . . .	126
6.2.5	SSQ . . . . .	127
6.3	Discussion . . . . .	128
6.3.1	Spatial Awareness . . . . .	128
6.3.2	Subjective Ratings . . . . .	129
6.3.3	Movement Profile . . . . .	131
6.4	Limitations and Future Work . . . . .	131
6.5	Conclusion . . . . .	132
<b>7</b>	<b>Conclusion</b>	<b>134</b>
7.1	Project summaries . . . . .	134
7.1.1	360° media acquisition . . . . .	134
7.1.2	Object removal in 360° media . . . . .	134
7.1.3	Cinematic virtual reality . . . . .	135
7.1.4	Multi-view 360° media . . . . .	136
7.2	Future work . . . . .	137
7.3	The future of the field . . . . .	138

<b>Appendices</b>	<b>142</b>
<b>A List of Acronyms</b>	<b>142</b>
<b>B List of all 360° media</b>	<b>143</b>
<b>C User study one: questions and tasks for all metrics</b>	<b>146</b>
C.1 Pre-experiment questionnaire . . . . .	146
C.2 DOCUMENTARY stimulus . . . . .	147
C.3 HORROR stimulus . . . . .	149
C.4 NARRATIVE stimulus . . . . .	150
C.5 Final questions . . . . .	151
<b>Bibliography</b>	<b>152</b>

# List of Figures

2.1	Use of seam carving to change an image's aspect ratio while avoiding distortion of important content. . . . .	24
2.2	Commercial HMDs . . . . .	27
2.3	The BBC's Surround Video. Figure from [1]. . . . .	31
2.4	MIT's Infinity-by-Nine. Figure from [2]. . . . .	32
2.5	Philips Ambilight TV. Figure from [3]. . . . .	32
2.6	Microsoft Research's IllumiRoom. Figure from [4]. . . . .	33
2.7	Shader lamps used to alter the perceived reflectance properties and create the illusion of shadows on a white model. Figure from [5]. . . . .	34
3.1	Filming with the Point Grey Ladybug3 on the roof of OnEustonSquare .	48
3.2	Point Grey's Ladybug3 connected to 17-inch MacBook Pro via Firewire 800 . . . . .	49
3.3	Frame from test documentary shot with Point Grey's Ladybug3 camera.	50
3.4	Frame from The Resistance of Honey. . . . .	51
3.5	GoPro array using a Freedom360 mount. . . . .	51
3.6	Examples of 360° video made available by the BBC and Peter Boyd Maclean. . . . .	52
3.7	Matterport's 360° RGB-D camera . . . . .	54
3.8	An example 3D model from Matterport . . . . .	54
4.1	FOV expansion and Graphcut Textures . . . . .	57
4.2	Rectilinear views of image 4.1a following HFOV expansion and region removal. . . . .	58
4.3	Rectilinear views of equirectangular video frames, showing object removal using the proposed graph cuts method. . . . .	59

4.4	Rectilinear views of the circular distortion introduced by the proposed graph cuts method at the south pole. . . . .	60
4.5	Rectilinear views of the tops of buildings, showing distortion introduced at the north pole by the proposed graph cuts method. Note that the removed section is behind the viewer, so cannot be seen in these images. . . . .	60
4.6	Failure cases of the graph cuts method: rectilinear views before and after cuts. . . . .	61
4.7	Seam carving can be used to preserve salient content during FOV expansion. Note that the removed section is behind the viewer, so cannot be seen in images b–d. . . . .	63
4.8	Grid shows non-homogeneous stretching used to preserve salient content. In these images, the stretched sections have been darkened to highlight areas that were altered. . . . .	64
4.9	Tripod removal using FOV expansion and graph cuts. . . . .	66
4.10	Inpainting directly in equirectangular. . . . .	67
4.11	Inpainting in equirectangular following rotation of south pole to the equator, as in Figures 4.9a and 4.9b. . . . .	68
4.12	Failure case of inpainting in equirectangular: large hole at south pole inpainted with geometric texture. . . . .	69
4.13	Inpainting of large hole at south pole. . . . .	70
4.14	Failure cases of inpainting method. . . . .	72
4.15	100° FOV rectilinear views of an inpainted video. . . . .	73
5.1	A 360° video being watched on a head-mounted display (left), a TV (right), and our SurroundVideo+ system (centre) . . . . .	79
5.2	Object placement task for HORROR stimulus. . . . .	84
5.3	Object placement task ensemble results. . . . .	88
5.4	Object placement task results by stimulus. . . . .	88
5.5	Boxplot of incidental memory results. . . . .	90
5.6	Boxplot of ensemble results from the MNEQ. . . . .	91
5.7	Results for the MNEQ by stimulus. . . . .	91

5.8	Boxplots for the four MNEQ sub-scales for the HORROR stimulus. . . .	92
5.9	Boxplots for the four MNEQ sub-scales for the NARRATIVE stimulus. . .	93
5.10	Boxplots for video enjoyment results. . . . .	94
5.11	Boxplots for display enjoyment results. . . . .	95
5.12	Boxplot of ensemble attention results. . . . .	96
5.13	Attention results by stimulus. . . . .	96
5.14	Boxplots for users' concern about missing something. . . . .	97
5.15	Boxplot of results for fear during horror. . . . .	98
5.16	Boxplot of SSQ total severity score. . . . .	99
6.1	A map of the temple stimulus. . . . .	107
6.2	The 3D model transition. . . . .	109
6.3	Frames from the Möbius transition. . . . .	111
6.4	Floating spheres represent camera locations the user can move to. . . .	114
6.5	An example pointing task. . . . .	116
6.6	Histograms of all pointing task results, including the angle's sign, for each transition type. . . . .	119
6.7	All participant answers for all pointing tasks. . . . .	120
6.8	Histograms of mean pointing task results for each transition type. . . .	121
6.9	Boxplot of average pointing task results for each transition type. . . .	122
6.10	Boxplot of average times between fading to grid environment and com- pletion of pointing task. . . . .	123
6.11	Boxplots of subjective ratings. . . . .	124
6.12	Boxplot of average time before initiation of the next transition during the returning phase. . . . .	127
7.1	Lytro's Immerge light-field camera. . . . .	140
7.2	Microsoft's free-viewpoint video capture system. . . . .	141
C.1	Object placement task for HORROR stimulus. . . . .	150
C.2	Object placement task for NARRATIVE stimulus. . . . .	151

## Chapter 1

# Introduction

Panoramic imagery has existed for centuries. Illusionism – an artistic technique in which a painting gives the illusion of a real scene – was frequently used in murals to give the impression the viewer had been transported to a different physical space [6]. Examples of such techniques being used to create sweeping views date back thousands of years. During the 19th century, a popular pastime was to view painted panoramas. By displaying these panoramas in purpose-built rotundas, the imagery surrounded the viewer, transporting them virtually to distant places [7].

A subcategory of panoramic media is 360° media. The term indicates that the horizontal field-of-view is a full 360°, surrounding the viewer horizontally. While the vertical field-of-view is not defined, it is often 180°, meaning the media covers a full sphere around the viewer. While it has been possible to paint 360° images for centuries, 360° photography and film are comparatively recent developments. Due to the physical limitations of pin-hole and lens-based cameras, it is not possible to capture fully spherical views using a single camera. Soon after the birth of photography, however, it was realised that by taking multiple images side by side, a single image could be cut together that captured a much wider field of view than was possible with a single camera alone. The Daguerreotype, an early form of photography, was introduced in the 1830s [8]. Examples of panoramas being created by combining multiple contiguous Daguerreotype images appeared around 10 years later [9].

Advances in computational photography now provide a mechanism called stitching, in which multiple views can be plausibly blended together into a single cohesive panorama in software. While digitally stitching views together to create panoramic media has existed since the 1990s [10], a recent convergence of technologies has cre-

ated a surge of interest in the medium. The high volumes of data required have benefited from advances in storage and transfer technologies, while the falling cost of high-quality cameras has reduced the barriers to entry for producers. Support for spherical video has been integrated into most popular video sharing sites, including YouTube, Facebook and Vimeo. Data released by these sites give an indication of the scale of consumption. In March 2017, more than one million 360° videos were available on the Facebook platform [11], while 2017 also saw the first 360° video on YouTube pass 10 million views [12].

The recent proliferation of virtual reality (VR) displays has also played a role in the rise of 360° media, as these devices allow the content to be consumed immersively. Orientation tracking provided by head-mounted displays (HMDs) allows the wearer to look around naturally inside a panorama. These affordances can illicit a powerful sensation of being in the space depicted in the panorama. Evidence suggests that substantial numbers of people are watching 360° videos immersively using HMDs. In January 2017, Samsung said 5 million Gear VR HMDs were in consumer hands globally, and that 10 million hours of video had been watched on these devices [13].

## 1.1 Research problem

Although 360° media has become extremely popular in a number of contexts, content production still presents substantial challenges. This includes all stages of the production pipeline. Planning for 360° videos is complicated by a lack of visual grammar for storytelling in this new medium. Capture of all forms of 360° media is complicated by the limitations of the current generation of camera hardware. Editing and post-production are hampered by a lack of tooling suitable for 360° media, while display is complicated by the wide variety of hardware and physical setups that consumers have.

Additionally, there is little academic work that looks at the evaluation of 360° media experiences. As a result, there are no agreed metrics by which these experiences can be investigated, and little data about the end-user experience.

The aims of this work are to:

1. Improve the production pipeline for 360° media.
2. Establish suitable evaluation techniques for 360° media experiences.

3. Use these techniques to evaluate 360° media in a number of contexts.

## 1.2 Scope

The production pipeline for 360° media is complex, involving many hardware and software components. Over the course of this project, the 360° media industry has seen substantial progress on a number of fronts, from the launch of a wide range of cameras purpose built for capturing of 360° video, to the integration of 360° media tools into popular post-production software. During this period, work was done to improve the content production pipeline. Specifically, an investigation of object removal in 360° media was performed, as this was an area lacking rigorous academic investigation at the time.

As popular video sharing websites began to support 360° video, acquisition of high-quality content became substantially easier. As a result, later work was able to focus on experience evaluation rather than content creation. There are a huge number of different types of 360° media. This includes both images and video. 360° videos come in a vast array of genres, such as comedy and mystery, but also journalism, travel and training videos. The work presented in this thesis focuses on broadly applicable evaluation techniques, with an emphasis on areas in which 360° media is likely to offer an improved experience or be at a disadvantage over traditional format media.

The work presented in this thesis looks specifically at 360° media that has been captured using a single camera or an array of cameras. It is important to note that 360° media can also be animated or computer generated, and that these mediums may impact the experience in various ways. Additionally, while 360° media supports only the three degrees of freedom associated with orientation, computer-generated media may potentially support positional tracking as well, which would likely also have an impact on the experience. The work presented in this thesis does not cover these scenarios, although it is hoped that many of the fundamental principles of this work could be applied in future investigations of the broader spectrum of media experiences.

As well as covering both images and video, 360° media can also be captured from a single viewpoint or multiple viewpoints. In the case of multi-viewpoint 360° media (MV360M) in which the user can choose to move from one camera location to another, a visual effect is required to transition between fixed camera locations. Such

experiences have been available for some time, from QuickTime VR in the 1990s [14] to the extremely popular Google Street View [15]. This work also aimed to explore the impact of the transition type when exploring MV360M in an immersive context. The work presented in this thesis considers only image and video content captured using a static camera, which is the most common technique for 360° media that is intended to be consumed immersively.

## 1.3 Contributions

### 1. Methodological contributions

- (a) Critique of evaluation techniques for passive, immersive media experiences.
- (b) Advice on the impact of the physical setup of 360° media experiences, particularly the impact of the chair type in use.

### 2. Substantive contributions

- (a) Research findings that address the impact of the display type in video-based 360° media experiences.
- (b) Research findings that address the impact of the transition type when exploring spaces in image-based multi-view 360° media.

### 3. Technical contributions

- (a) Techniques for object removal in 360° images, with extensions for video in simple cases.
- (b) Consideration of Möbius transitions as a mechanism for scene traversal in multi-view 360° media.

## 1.4 Publications

The research that forms part of this thesis has led to several publications. These are:

Andrew MacQuarrie and Anthony Steed. Object removal in panoramic media. In *Proceedings of the 12th European Conference on Visual Media Production*, pages 2–11. ACM, 2015. [16]

Andrew MacQuarrie and Anthony Steed. Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video. In *Virtual Reality (VR), 2017 IEEE*, pages 45–54. IEEE, 2017. [17]

Andrew MacQuarrie and Anthony Steed. The effect of transition type in multi-view 360° media. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1564–1573, 2018. [18]

## 1.5 Structure

Chapter 2 provides an overview of the current status of the 360° media pipeline, and contextualises this in the broader area of immersive content. Existing techniques for editing standard format media are discussed, as well as the challenges of extending these techniques to work for panoramic media. Immersive displays are introduced and critiqued. A survey is then presented of work relevant to the evaluation of immersive 360° media experiences.

The remainder of the thesis is divided into two sections. This reflects the changing landscape of the 360° media field over the last four years. At the beginning of this project, high-quality 360° content was difficult to source. In order to investigate the evaluation of 360° media experiences, it was necessary to create suitable content. As a result, content was created in collaboration with BBC R&D. This effort is discussed in Chapter 3. During this content creation, work was done to improve the production pipeline. Specifically, object removal in 360° images was investigated, and extensions for video demonstrated in simple cases. Work on object removal in panoramic media is presented in Chapter 4.

Over the duration of the project, an explosion of interest in 360° media occurred. In tandem with this, the tools for the production and distribution of content radically improved. High-quality content became readily available from numerous sources. As a result of this shift in the field, we were able to turn our attention to the evaluation of 360° media experiences, without needing to further develop our own content or the production pipeline. Content acquisition via these other sources is also discussed in Chapter 3.

In the second half of this thesis, evaluation of 360° media experiences is considered. Evaluation of single-view 360° video is discussed in Chapter 5. A between-group experiment with 63 participants was conducted. Participants watched several pieces of 360° video content in one of three different displays: a HMD, a projection-based focus-plus-context display, and a standard 16:9 TV. Multiple metrics were investigated, including spatial awareness, memory, narrative engagement, a viewer's concern about missing something, enjoyment, and the display's ability to guide attention. The results indicated that the HMD offered a significant benefit in terms of enjoyment and spatial awareness, while the focus-plus-context display offered a significant improvement in enjoyment over traditional TV. Drawing attention and a viewer's concern about missing something were also not significantly different between display conditions; these results were unexpected, and may indicate that using a fixed chair in a HMD experience places a soft limit on the field of regard.

Evaluation of image-based multi-view 360° media experiences is discussed in Chapter 6. A between-subject experiment with 31 participants is described, in which the effect of the transition type used to move users between camera locations in the media was investigated. Three transition types were examined: teleport, a linear move through a 3D model of the scene, and an image-based transition using a Möbius transformation. The metrics investigated included spatial awareness, users' movement profiles, transition preference and the subjective feeling of moving through the space. The results, which are discussed fully in Chapter 6, indicate that trade-offs between transitions will require content creators to think carefully about what aspects they consider to be most important when producing such experiences.

Chapter 7 concludes, and proposes future research based on the findings presented throughout the thesis. Chapter 7 also takes a broader look at the future of captured virtual reality content, and discusses how the work presented here is likely to apply in this rapidly changing field.

## Chapter 2

# Background

There are three main stages in the pipeline for filmed 360° content. First, the material must be captured using a single camera or array of cameras. Next, the material is manipulated and edited in such a way as to make it suitable for display. Finally, the content is displayed using a playback mechanism. In this chapter, literature relevant to these stages is reviewed, as well as previous work relevant to the evaluation of 360° media experiences.

In section 2.1, the capture of 360° media is discussed. Additionally, state-of-the-art techniques in traditional format media editing are examined, and issues with applying these techniques to 360° media identified.

The immersive display of panoramic media is presented in section 2.2. A range of immersive technologies have been developed over time. These different technologies have resulted in a range of VR displays, each with different affordances and limitations. These technologies are described, along with the resulting displays and their respective properties.

Section 2.3 contains a discussion of how to evaluate immersive display systems. When considering 360° video, the focus of this discussion is on areas where immersive displays are likely to offer some benefit over, or be at a disadvantage to, traditional format displays. When considering MV360M, the focus is on the likely impact of different transition types when exploring scenes captured in this medium. Issues related to comfort – in particular simulator sickness – are examined for both media formats.

## 2.1 Panoramic media

### 2.1.1 Panoramic media capture

Panoramic media refers to images and video that have a large field of view (FOV), covering up to a full-sphere around the viewer of  $360^\circ$  horizontally and  $180^\circ$  vertically. This is often referred to as the *viewing sphere*. While panoramic media is not a new concept, the field has seen a rapid increase in interest over the last few years. This rise has been driven by the falling cost and rising quality of capture and playback devices. Mass-produced, wide-angle “action cameras”, such as the GoPro HERO, have brought the cost of building a  $360^\circ$  camera rig down significantly, while an array of purpose-build devices are also entering the market [19]. Playback support has been integrated into YouTube, allowing easy viewing in browsers and phones [20]. Meanwhile, head-mounted displays such as the Oculus Rift and Google Cardboard have brought immersive, panoramic media playback into the mainstream.

Creating panoramas by combining several overlapping views is a technique that has been used in computational photography for around 20 years [21, 22]. The process involves finding a  $3 \times 3$  homographic transformation matrix for each image. These matrices can be used to warp the images into alignment. Corresponding feature points between images are used to find these homographies, and improvements in the methods of identifying these correspondences has produced automatic tools that are effective and simple to use [23]. Each view is then corrected for lens distortion and blended together using image processing [24]. Playback is then achieved using software or specialised hardware. Software viewers are available for desktops and mobile phones. For immersive viewing experiences, head-mounted displays can be used that allow the viewer to look in any direction by turning their head naturally.

### 2.1.2 Cinematic virtual reality

Cinematic virtual reality (CVR) is a broad term that could be considered to encompass a growing range of concepts, from passive  $360^\circ$  videos, to interactive narrative videos that allow the viewer to affect the story. Work in lightfield playback [25] and free-viewpoint video [26] means that soon viewers will be able to move around inside captured or pre-rendered media with six degrees of freedom. Real-time rendered, story-led experiences also straddle the boundary between film and virtual reality.

By far the majority of CVR experiences are currently passive 360° videos. In January 2017, Samsung said five million Gear VR HMDs were in consumer hands globally, and that 10 million hours of video had been watched on these devices [13]. In March 2017, more than one million 360° videos were available on the Facebook platform [11]. These videos span a broad spectrum of genres, from news and journalism to comedy and horror. In this thesis, the term CVR is used to mean monoscopic, passive, fixed-viewpoint 360° videos, as these are by far the most commonly available type of video for virtual reality.

### 2.1.3 Multi-view panoramic media

Panoramic media increases affordances over traditional formats by allowing the user to view a larger FOV. Such experiences are typically fixed-position, however, meaning that while the user can look around from a single location, they are unable to control the movement to another location. Multi-view 360° media (MV360M) can help to facilitate the exploration of a captured scene, by allowing users to actively view the space from multiple locations through their own control. This increase in agency could be considered to make a system more immersive, and may lead to more engaging experiences. Such systems can use 360° images or video. Various VR experiences have made use of these content types in different ways. QuickTime VR allowed the user to explore a scene by navigating between 360° images captured from different locations [14]. This type of content has become commonplace, for example in systems such as Google Street View [15].

Image-based MV360M benefits from requiring only one camera that can be moved between locations, as well as minimal data processing. Video-based MV360M has historically been less popular, as such productions typically require a camera at each location in order to capture an event. Additionally, stitching high-resolution video from multiple locations requires large amounts of data processing and bandwidth. The falling cost of cameras, and improvements in stitching technology [27], mean that video-based MV360M is becoming more common. Video-based MV360M has been live streamed from events including award shows [28] and concerts [29], allowing users to choose between views in real-time.

### 2.1.4 Panoramic media editing

Although it remains a practical challenge for filmographers, the capture of 360° media is now well understood. Editing panoramic media, however, brings new challenges. Due to the fact that filming without the need for post-production is often prohibitively difficult, the editing of standard format images and video has been studied extensively. This research has resulted in the development of many algorithms that produce excellent results in areas such as object removal, hole filling, retargeting and reshuffling. However, it is not yet clear what must be done to allow these algorithms to produce equally good results on panoramic media. Key differences – such as the projection that spherical content must undergo in order to be edited and stored effectively – strongly indicate that these algorithms cannot be applied without alteration. Additionally, there are likely improvements which can be made by adapting these techniques to make use of the properties of panoramic media, such as the wealth of additional information that is captured over and above that of a regular camera.

In this section the properties of various projections are discussed. Additionally, some of the above mentioned algorithms for media editing are introduced. Later, in Chapter 4, these components will be combined to perform editing of panoramic media. Specifically, object removal in 360° media will be explored.

#### 2.1.4.1 Projections

While image processing operations could be done on the sphere, it is often simpler for the editing and storage of panoramic media to project this sphere onto a plane. The problem of projecting a sphere onto a plane has been extensively studied for thousands of years, due to the need to create 2D maps of the earth [30]. Through the study of cartography, a large number of different types of map projection have been designed.

The projection of a sphere onto a plane inherently introduces distortion. As a result, each type of map projection seeks to preserve some aspect of the original content. The simplest projection is the equirectangular projection. This projection maps the lines of longitude to vertical straight lines and the lines of latitude to horizontal straight lines [31]. A special case of the equirectangular projection is the plate carrée projection, in which the equator is the undistorted latitude. As the plate carrée is very common, it is usually referred to simply as “equirectangular” – a convention also fol-

lowed in this document. In this projection, the x coordinate of the image maps directly to the longitude and the y coordinate maps directly to the latitude. As a full-sphere panorama has a horizontal FOV (HFOV) of  $360^\circ$  and a vertical FOV (VFOV) of  $180^\circ$ , the equirectangular projection results in a rectangular image with a 2:1 aspect ratio.

Due to the nature of the equirectangular projection, there is very little distortion around the equator but substantial distortion at the poles. Equirectangular projections are not often used in cartography because of this large pole distortion. However, due to their simplicity and the ease with which scenes can be understood, equirectangular is probably the most commonly used projection in panoramic media. Other notable projections include: Mercator, which preserves shapes locally; Sinusoidal, which preserves relative areas; and rectilinear (perspective), which preserves straight lines but introduces stretching distortions at the edge of the image when covering a FOV above  $40^\circ$  [32].

#### 2.1.4.2 Graphcut Textures

In Chapter 4 it will become necessary to cut two images together in a way that disguises a join. Combining two images together in a plausible way can be achieved using a technique known as Graphcut Textures [33]. Two images are overlaid at their desired positions. A cut between them is then identified with the intention of disguising the join. This cut is found using a min-cut/max-flow optimisation.

To perform the min-cut/max-flow optimisation, a graph is constructed. Each pixel is connected to its neighbours, with weights favouring cuts at areas in the original images that appear similar. The weight for the arc between adjacent pixels  $s$  and  $t$  is defined by examining the surrounding pixels in the two images, using the formula:

$$M(s, t, \mathbf{A}, \mathbf{B}) = \|\mathbf{A}(s) - \mathbf{B}(s)\| + \|\mathbf{A}(t) - \mathbf{B}(t)\|$$

where  $\mathbf{A}(s)$  is the patch from image  $\mathbf{A}$  at pixel  $s$ ,  $\mathbf{B}(t)$  is the patch from image  $\mathbf{B}$  at pixel  $t$ , and  $\|\cdot\|$  denotes an appropriate norm. Arcs with infinite weight (“*constraint arcs*”) are used to ensure certain pixels are taken from a specific image. For example, for images  $\mathbf{A}$  and  $\mathbf{B}$  placed side by side and being joined by a vertical cut, pixels at the left hand side and right hand side of the overlapping area are taken from image  $\mathbf{A}$  and  $\mathbf{B}$  respectively. Min-cut/max-flow is then used to find an optimal seam. This seam identifies which image each of the unconstrained pixels should be copied from. Pixels

to the left of the cut are taken from image **A** while pixels to the right of the cut are taken from image **B**. This algorithm can produce compelling results, particularly in scenes with similar areas in both images that allow a good cut.

#### 2.1.4.3 Retargeting

As will be shown in Chapter 4, it can be useful to alter the FOV of a displayed panorama. This can be different from the FOV a panorama captures. For example, a panorama that captures 360° HFOV could be warped to be displayed over only 180°. This can be achieved by altering the height and width of some projections, including equirectangular. Altering the width and height of media is referred to here as *retargeting*.

Due to the heterogeneous nature of display devices, retargeting media is a well studied area. The simplest method is to resize the content equally using a scaling algorithm such as bicubic interpolation [34]. Algorithms that aim to improve on this generally attempt to minimise noticeable distortion to salient content.

Seam carving is a method to retarget images. It does this by removing or adding vertical or horizontal connected pixel-wide cuts in areas where the change will be least noticed. For example to reduce the size of the image on the horizontal axis, vertical seams are removed in areas with little energy as defined by some energy function [35]. The energy function could encode different saliency measures, from pixel gradients to facial detection. Seams are removed iteratively in a greedy way until the image conforms to the desired dimension, an example of which can be seen in Figure 2.1. The results can be very impressive, especially in images where mild deformations are not jarring, such as mountain ranges. In images such as crowds of people, that have few low-energy areas and features such as faces that cannot be deformed, the results can be poor.

The position of the seam is based on an energy function. One method is to look at the areas of the image that have small amounts of energy. This is called “backward energy”. Another method called “Forward Energy” was proposed by Rubinstein et al. in their paper “Improved seam carving for video retargeting” [36]. Instead of choosing an existing area with little energy to create a seam, they instead look at the energy introduced by adding or removing a seam. It is claimed that forward energy produces



(a) Original image with one vertical and one horizontal seam shown in red



(b) Aspect ratio changed using traditional scaling



(c) Aspect ratio changed using seam carving

**Figure 2.1:** Use of seam carving to change an image’s aspect ratio while avoiding distortion of important content. Figure from [35].

better results than backwards energy.

Seam carving has also been extended to video in a variety of ways. Rubinstein et al. proposed that a 2D seam manifold be found that cuts through the 3D space-time volume of the video [36]. As the dynamic programming approach used in image seam carving is not suitable for this 3D problem, it is replaced with graph cuts. As the seam is continuous throughout the space-time volume a coherent video can be produced. This comes at the expensive of speed, as the graph cuts solution is computationally expensive to find. Additionally, a cut through the entire video must be found at once.

An alternative approach was suggested Grundmann et al. in “Discontinuous seam-carving for video retargeting” [37]. Each frame is constrained to be similar to the previous frame, but the seams can be temporally discontinuous. Capable of working at two fps for 400x300 video, this algorithm is around four times faster than that proposed

by Rubinstein et al. [36]. As the algorithm considers frames sequentially it can also work on streaming video.

Other retargeting methods include non-homogeneous warping [38], and shift-maps which discretize the problem into the rearrangement of pixel positions [39].

#### 2.1.4.4 Inpainting

In Chapter 4, we will investigate inpainting in panoramic content. Inpainting is the filling of holes in a plausible way. It is a powerful technique that has gained extensive popularity. While other forms of inpainting such as Shift-maps have been proposed [39], the technique most generally used is a patch-based method derived from previous work on texture synthesis [40, 41]. Holes are filled from the outside in, propagating structure and texture by copying suitable content from elsewhere in the image or video, using a patch-based similarity measure to find a nearest-neighbour field (NNF).

The construction of the NNF presents complexity issues if approached in a brute force fashion. However, several excellent improvements have been proposed that calculate approximate nearest-neighbour fields (ANNF) at a fraction of the complexity. PatchMatch was a seminal work in this area [42]. It uses the coherence of images to propagate good matches, found via random sampling, to a pixel's neighbours.

An excellent inpainting implementation is Adobe Photoshop's content-aware fill tool. For the inpainting step, content-aware fill uses a patch-based hole filling method based on Space-time Video Completion, originally proposed by Wexler et al. [43, 44]. PatchMatch is used to create the ANNF, which allows interactive speeds to be achieved.

#### 2.1.4.5 Working with panoramic media

Sacht et al. have done work in panoramic media in the context of straight line and face detection [45]. They reached the conclusion that working in panoramas requires finding a suitable projection for the job at hand. They discovered that the local shape preservation properties of the Mercator projection facilitates the identification of faces, while the perspective projection is required for line detection.

Inpainting in 360° images was considered in work parallel to our own by Zhu et al. [46]. They proposed a structure-rectifying warp to support image completion. In contrast to their work, in Chapter 4 we consider scenarios in which simple projections are sufficient for object removal, and discuss when more complex strategies are re-

quired. Additionally, we propose a further technique for object removal, and consider extensions for video in simple cases.

## 2.2 Immersive displays

### 2.2.1 Head-mounted displays

There are several different types of head-mounted display (HMD). While there are a growing number of HMDs that support augmented reality, the vast majority of HMDs are currently virtual reality devices in which the real-world is entirely occluded. The current generation of popular VR HMDs all follow a similar fundamental design. A video screen, or pair of screens, is fixed directly in front of the eyes by some form of head-worn mount. This screen is typically only centimetres away from the eyes. As the human eye would struggle to focus on such a close surface, lenses are used to project the screen to a more comfortable accommodation distance. For example, the optics in the Oculus Rift CV1 “are equivalent to looking at a screen approximately 1.3 meters away” [47].

#### 2.2.1.1 Commercialisation

HMDs have been in use in academia for decades. A recent convergence in display, tracking and processing technologies mean that HMDs are now reaching a quality and form factor that consumers find acceptable. As the displays and sensors used in many modern HMDs are mass produced components originally designed for mobile phones, the cost to create a HMD is orders of magnitude smaller than it would have been even a decade ago. As a result, the price of HMDs is now entering a range suitable for the consumer market.

Consumer HMDs cover a broad spectrum of cost and quality. Some lower-end HMDs allow users to slot in their mobile phone, utilising its screen and processing capabilities. At the lowest end of the spectrum is the Google Cardboard, shown in Figure 2.2a<sup>1</sup>. Costing around £3, these devices are composed of two small lenses, held inside a cardboard mount. Any modern smartphone can be placed inside. The phone’s inertial measurement unit (IMU) is used to provide orientation tracking. While cheap, the quality of such experiences is relatively poor. This is due to poor build

---

<sup>1</sup>Creative Commons image by othree, available from [https://commons.wikimedia.org/wiki/File:Assembled\\_Google\\_Cardboard\\_VR\\_mount.jpg](https://commons.wikimedia.org/wiki/File:Assembled_Google_Cardboard_VR_mount.jpg)



**Figure 2.2:** Commercial HMDs

and lens quality, the restricted processing capabilities, as well as the limitations of IMU tracking, which are discussed in section 2.2.1.2 below. Additionally, these devices have limited input mechanisms.

A step up from the Cardboard is the Gear VR HMD, shown in Figure 2.2b. This device makes use of the display and processing power of a Samsung phone that is slotted into it. In comparison to the Cardboard, the Gear VR is more sturdily built using moulded plastic, more comfortable to wear, has increased input options via a built-in touchpad, has higher quality lens optics, and contains its own gyroscope for improved tracking [48]. As a result of these characteristics, the Gear VR has become extremely popular, shipping five million headsets by January 2017 [49]. Similarly to the Cardboard, however, they only provide orientation tracking.

At the higher end of commercial HMDs are devices like the Oculus Rift CV1, shown in Figure 2.2c, and the HTC Vive. These tethered HMDs require high-end PCs with modern graphics cards. Although they differ in their optical tracking technology, both provide tracking that allows users to walk around with high accuracy and low latency. Both also can be used with tracked input controllers held in each hand, opening up a wealth of possible interaction techniques.

### 2.2.1.2 Tracking

When a viewer moves their head, the virtual world depicted in the HMD must update accordingly in order for the illusion of reality to be maintained. To allow this, HMDs must be tracked.

Tracking the three axes of rotation that constitute the direction the viewer is looking, known as orientation tracking, is a basic requirement for a HMD. Each of these three axes of rotation are considered a degree of freedom (DoF) that the HMD af-

fords, and as a result HMDs that provide only orientation tracking are considered 3DoF HMDs. Orientation tracking can be performed using IMUs [50]. IMUs are composed of accelerometers, which measure proper acceleration, and gyroscopes, which measure orientation. IMUs are very low latency, so are ideal for updating the screens of HMDs operating at high frequency rates. It has been shown that the delay for a HMD to respond to head motion is a contributing factor to simulator sickness [51], and indeed Oculus CTO John Carmack has stated this latency is “one of the primary causes of simulator sickness” [52]. As a result, IMUs have become commonplace in HMDs. IMUs suffer from accumulated errors, however, which lead to drift over time. For 3DoF HMDs tracked purely through IMUs, this would present itself as the virtual world rotating slowly over time on the horizontal axis.

In order to correct for drift, and accurately provide positional data, an additional type of tracking is required [50]. In popular commercial HMDs, this is achieved using some form of optical tracking, with the two tracking methods used in tandem through sensor fusion. Optical tracking can be outside-in, in which the device is tracked by an external sensor facing the HMD, or inside-out, in which the HMD tracks itself through outward facing sensors [50]. For example, the Oculus Rift CV1 uses outside-in tracking, with an infrared USB camera being used to detect infrared LEDs placed inside the case of the HMD unit. While effective, this method has limitations on the size of the tracking volume defined by the line-of-sight of the tracking sensor. Inside-out tracking, for example using RGB cameras on the HMD itself, is an increasingly popular form of tracking that is not bound to a tracking volume, and a number of new HMDs are adopting this technique [53]. HMDs that provide both orientation and positional tracking are said to have 6 degrees of freedom (6DoF).

### 2.2.1.3 HMD characteristics

There are a number of characteristics that define how “good” a HMD is. The resolution of the device is defined by the screen and optical system. The current generation of HMDs, while improving, still do not come close to rendering at a resolution in which individual pixels are not clearly perceptible to the fovea of the human eye [54]. This is partially a factor of the resolution of the screen in use, but is also compounded by the proximity of the screen to the eye, and the magnifying effect of the lenses. This

impacts the experience in a number of ways, and creates issues around interface design, e.g. small text may not be legible.

Another factor is the field-of-view (FOV) supported by the device. Again this is a factor of the size of the screen, and the optical array in use. Humans have a horizontal FOV of around  $180^\circ$  without moving the eyes, and about  $290^\circ$  including eye rotation [55]. Despite this, the biggest selling devices in the current generation of commercial HMDs, such as the Oculus Rift CV1 and the HTC Vive, provide a horizontal FOV of approximately  $90^\circ$ .

The processing power available may or may not be a characteristic of the HMD itself, depending on the type of device in question. HMDs are either stand-alone, meaning the processing is done on the HMD itself, or tethered, meaning they are connected to a PC on which the processing takes place. Generally speaking, modern tethered HMDs are far more powerful than stand-alone HMDs, as they can make use of fully fledged graphics cards that require large amounts of power and ventilation to run.

### 2.2.2 Projection-based displays

Another way to create immersive visual experiences is to use projection. This allows the creation of displays that can be larger than would be practical using fixed screen-based technologies, and do not fully occlude the real world as is the case for VR HMDs. A projection-based display was used in the experiment described in Chapter 5. One of the simplest such system is the CAVE<sup>TM</sup>-like display.

#### 2.2.2.1 CAVE<sup>TM</sup>-like displays

One way to create almost perfect wide-field-of-view projected visuals is to use a CAVE<sup>TM</sup>-like display [56]. Such displays use flat white walls arranged in a cube around the viewer. The even colour of the walls means that the surface has ideal radiometric qualities for projection, while the planar shape of the surfaces mean complex computations are not required to establish the correct projection needed to achieve the desired result. To provide perspective-correct rendering, head tracking can be used.

One important property that CAVE<sup>TM</sup>-like displays often possess is that their field-of-view can match that of human vision. While it is possible to have all six sides of the cube available, practical limitations mean these displays generally have fewer sides. Assuming a cube with three walls, with a user positioned centrally within the space and

looking away from the missing wall, the entire horizontal field-of-view of the user can be covered.

#### 2.2.2.2 Focus+Context displays

In Chapter 5, a projection-based Focus+Context display will be used. In this section, these displays are introduced and the details of creating such systems examined.

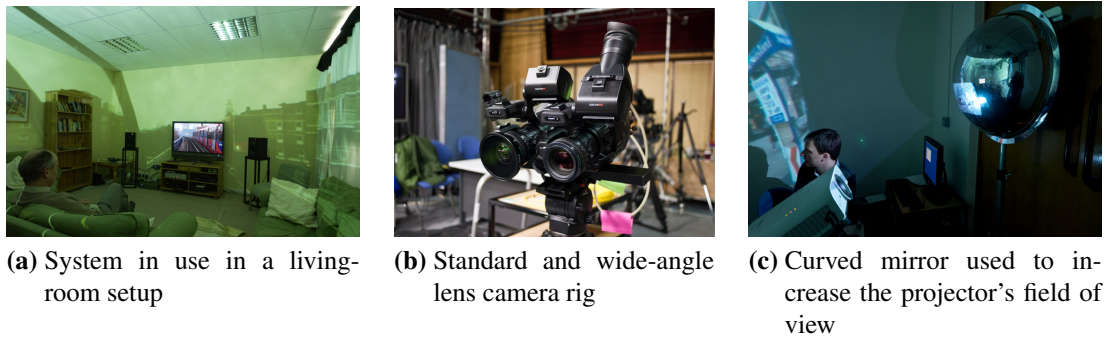
#### **Surround Video**

The BBC's Surround Video was created in 2010 as an experiment in producing a more immersive experience while watching video content [1]. The field of view was increased by projecting content around a standard HD TV onto the walls and furniture of the room, as can be seen in Figure 2.3a. While the surround content was aligned with the TV content, no projection mapping or radiometric compensation was employed. The test room contained little furniture, and it was assumed that the room had reasonable radiometric and geometric properties for projection, i.e., white in colour and reasonably flat.

To produce video for this setup, content was filmed simultaneously using a standard lens and a wide-angle lens, as shown in Figure 2.3b. During playback, footage shot using the wide-angle lens was projected off of a curved mirror to create an extremely large field of view. This setup is shown in Figure 2.3c. While the field of view was large, the resolution, brightness and focus of the images suffered as a result of this projection method. As it was expected that the surround content would be viewed almost exclusively in peripheral vision, it was hoped the reduced resolution of the human eye away from the fovea would mitigate these issues. However, during tests, it was discovered that users often looked directly at the peripherally projected content. In fact, the director of “Broken”, the short film commissioned to showcase the Surround Video's potential, at times put important visual content in the surrounding projections, thereby encouraging users to look directly at it.

#### **Infinity-by-nine**

Producing content for the BBC's Surround Video presented many technical challenges [1]. A possible alternative to capturing surround content was considered by the MIT Media Lab in their Infinity-by-Nine system [2]. This system analysed standard video content to generate surround visuals, which were projected onto flat white



**Figure 2.3:** The BBC's Surround Video. Figure from [1].

screens placed to either side a TV as shown in Figure 2.4. The projected visuals attempted to capture the general impression of the content on screen, for example by matching the colours or extending the horizon. Although the peripherally projected content was not as context-specific or detailed as that of the Surround Video, the fact that the experience could be created using pre-existing content was a major benefit.

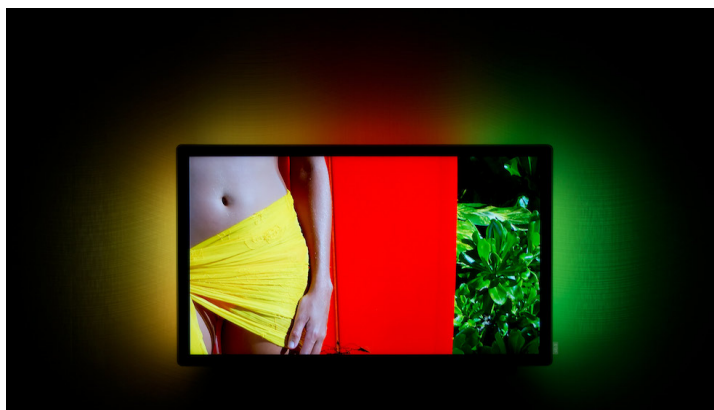
This concept was commercialised to some extent by Philips in their Ambilight TV [3]. This TV analyses the displayed content and projects light onto the surface behind and around the TV, as shown in Figure 2.5. The colour and intensity of the projected light is designed to enhance the viewing experience and increase the immersive properties of the display. However, the light projected onto the surrounding area had very limited resolution as it was produced by a single strip of LEDs around the edge of the TV's rear casing. The concept was expanded in a research project by Philips, in which several LED strips were placed around the room allowing a more immersive experience to be created [57].

### **IllumiRoom**

In 2013 Microsoft Research produced the IllumiRoom [4]. This system consisted of peripheral projections centred around an HD TV in a setup similar to the BBC's Surround Video, as shown in Figure 2.6. The peripheral projections relied heavily on projection mapping techniques, which are discussed in section 2.2.2.3. The major difference between the concepts of IllumiRoom and Surround Video is that, while Surround Video used purely filmed content, the IllumiRoom's projections were rendered on the fly as an extension of the video game content that was being rendered for the TV. By aligning the projections to the geometry of the room, compelling visual effects



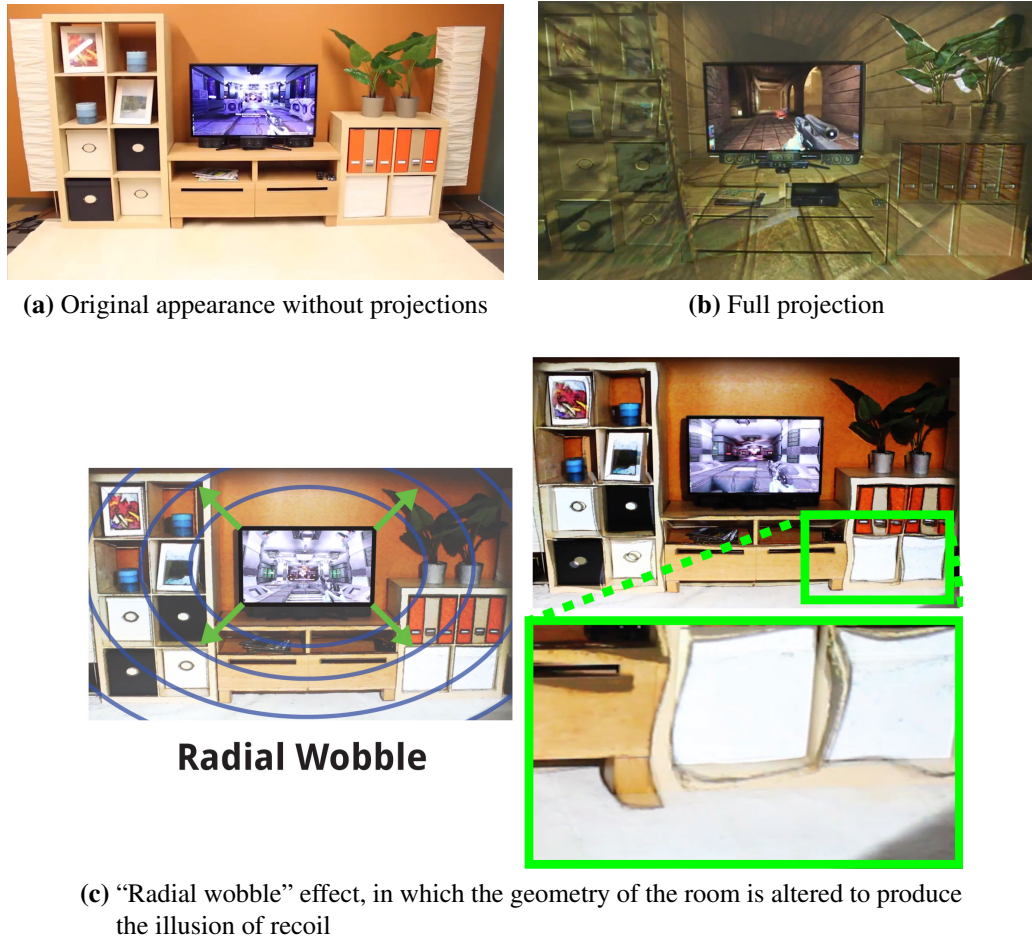
**Figure 2.4:** MIT's Infinity-by-Nine. Figure from [2].



**Figure 2.5:** Philips Ambilight TV. Figure from [3].

can be produced. These projection mapping techniques are discussed further in section 2.2.2.3. The Microsoft team demonstrated effects such as: altering the tone of the room to match the content better; relighting the room using dynamic shadows to give the appearance of street lights rushing by during a racing game; and altering the geometry of the room slightly to give the illusion of recoil during a first-person shooter game, shown in Figure 2.6c.

In 2014 Microsoft Research extended the IllumiRoom concept further to create a system called RoomAlive [58]. RoomAlive used six projectors – each paired with a Microsoft Kinect – to spatially augment an entire room. In this system, the HD TV around which the IllumiRoom system had centred was removed. Users were tracked in the space using the Kinect sensors, allowing the user to interact in interesting ways. For example, a game of whack-a-mole was created in which moles appeared anywhere

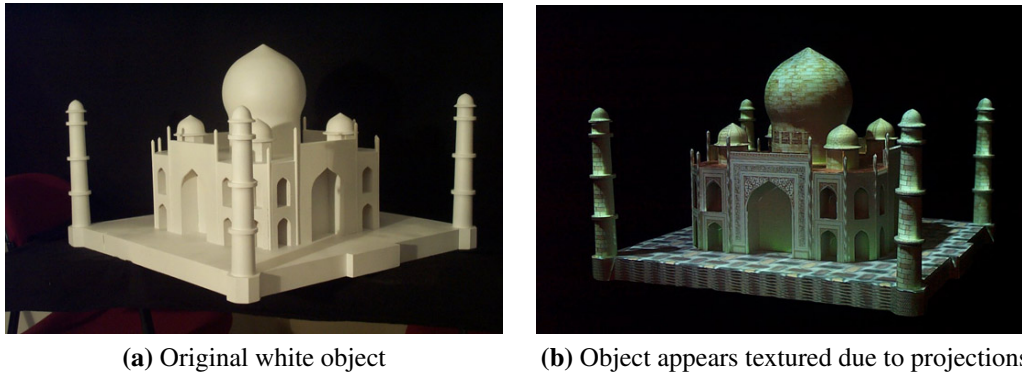


**Figure 2.6:** Microsoft Research's IllumiRoom. Figure from [4].

in the room, requiring the user to hit them with their hands or feet. Additionally, head tracking provided by the Kinects allowed for perspective correct rendering, providing a CAVE™-like experience [56]. While RoomAlive provides a significant contribution in terms of calibrating multiple projector/camera units, the visual impact of displayed content was hampered by a lack of radiometric compensation and blending between multiple projectors.

### 2.2.2.3 Projection mapping

Spatial augmented reality (SAR) – otherwise known as ‘projection mapping’ – has been in use for decades. The concept involves projecting onto a non-planar surface, often aligned such that the non-planarity enhances the visuals. The concept originated in the entertainment industry, when in 1969 Disneyland used the technique in their Haunted Mansion ride [59]. On this ride, white busts were animated with projected videos of faces, aligned in such a way as to give the illusion that the bust itself was in motion. The



**Figure 2.7:** Shader lamps used to alter the perceived reflectance properties and create the illusion of shadows on a white model. Figure from [5].

three dimensionality of the display surface created a more compelling illusion than a flat display surface, as the viewer experienced some 3D cues such as binocular disparity and motion parallax. This technique was described in academic work for the first time in 1998, when Raskar et al. presented their vision for “The Office of the Future” [60].

Since then projection mapping has steadily gained popularity [59]. Some of the most impressive results have come from promotional material, such as a viral campaign built by Marshmallow Laser Feast for the launch of the Sony PlayStation Video store [61] and Bot and Dolly’s “Box” [62]. A large part of what makes these projects so impressive, however, is that they rely in part on projected visuals that will work only when viewed from a specific viewpoint.

One technique to project onto complex geometries is to build a 3D model of the space and calculate automatically what needs to be projected to augment the real-world objects. This approach was considered by Raskar et al. in their paper “Shader lamps: Animating real objects with image-based illumination” [5]. The idea is to recreate the appearance of a textured object by replacing it with a neutral object of the same shape and using projected images to produce a similar reflectance, as shown in Figure 2.7.

Many elements affect the achievable results in a projector setup. There have been a number of good investigations into these issues [5, 63, 4]. One of the most critical aspects that will affect the final result is the attributes of the projector in use, such as brightness, resolution etc., and the reflectance properties of the projection surface.

In some cases the true geometry and reflectance properties of a surface can be overcome to produce the desired result, completely negating the real world object. This

process is called radiometric compensation and involves projecting light that, when reflected off of the object, will produce the desired colour and intensity [64]. This is achieved by capturing the surface geometry and reflectance properties, as well as other light present in the physical scene, by using a projector-camera setup. The results achieved by these techniques can be impressive. Radiometric compensation, however, can only alter the appearance of a surface so far. For example, it is not possible to make a red surface appear blue, as the surface itself will absorb the required light wavelengths [4].

## **2.3 Evaluation**

Evaluating a system can help researchers to understand the underlying psychological processes involved when that system is in use. The focus of this discussion is on the evaluation of VR systems in the context of experiences based on panoramic media. A number of evaluation techniques are discussed, highlighting their appropriateness for evaluating immersive display systems in relevant contexts.

### **2.3.1 Evaluation of cinematic virtual reality experiences**

In Chapter 5, a user study will be described that investigates CVR. Evaluation of CVR experiences presents several issues. While immersive displays have been evaluated for many years, elements inherent in CVR make adapting existing techniques challenging. Evaluation of virtual environment (VE) displays generally focuses on task completion (for a survey, see Bowman et al. [65]). This presents a particular problem for CVR evaluation, as these experiences are generally passive and therefore do not support the types of task that are purely observational in practice. Likewise, evaluation techniques common for standard format passive displays may lack the ability to assess the impact of highly immersive experiences. Although CVR experiences are considered “passive”, the viewer can look around and is still engaging in multiple cognitive activities. These activities may or may not be supported by the display configuration, such as the field of view (FOV), light levels, etc.

#### **2.3.1.1 Evaluating passive experiences in immersive displays**

Philips evaluated an early version of the Ambilight TV [66]. This repeated-measure experiment, in which participants rated criteria such as presence and naturalness on a five-

point numeric scale, was informed by the International Telecommunication Union's BT500 methodology [67]. While the BT500 is a common tool for measuring subjective perception of image quality, it is not clear that it can be extended to immersive experiences and concepts such as presence. While significant results were found, the repeated-measure design may be susceptible to demand characteristics, as participants can easily guess the hypotheses under test.

Further research by Philips for an Advanced Ambilight system included objective measures via physiological monitoring, such as heart rate, skin conductance and respiration [57]. Using physiological monitoring can be challenging, however, and significant results between conditions were not obtained for these measures. Heart rate and skin conductance responses were used successfully by Reeves et al. in their investigation of the effects of screen size on arousal and attention [68]. Their technique for attention requires content with frequent cuts in order to trigger orienting responses, however, and their technique for arousal is best suited for arousing content (e.g. videos containing violence and sex).

A review of literature relevant to immersive display evaluation was completed by Schnall et al. in their investigation of fulldomes, the immersive dome-based projection displays most known for their use in planetariums [69]. With a focus on the educational benefits of fulldome displays, their review included research on immersion, presence, attention, memory and social factors. They presented several suggestions, including ensuring as much consistency as possible between display conditions to reduce the risk of confounding factors. They did not propose a framework for evaluating such displays, however, and did not conduct any experiments.

Fonseca and Kraus explored how the level of immersion of a display impacts the viewing experience when viewing 360° videos [70]. In a between-groups experiment, participants watched an emotional 360° video about the meat industry. Twenty-one participants watched the video in an Oculus Rift DK2 HMD, while 21 participants watched the same video on a 10.1-inch tablet. The results indicated that the more immersive condition significantly enhanced pro-environmental attitude, and increased participants' level of sympathy for the characters in the video.

### 2.3.1.2 Presence

Immersive displays have often been measured by the sense of presence they create. This document follows the convention of using the term “immersion” to mean an inherent characteristic of the display system in use, while “presence” denotes the viewer’s sensation of “being there” [71]. There have been many techniques suggested for the measurement of presence. These include subjective self-assessment questionnaires such as that proposed by Witmer and Singer [72], physiological monitoring [73], and realistic physical responses to situations (“response-as-if-real”) [74].

While often used, questionnaire-based methods of presence measurement have been criticised [75, 76]. Physiological monitoring has also been criticised as a measure for presence, and has shown limited ability to produce significant results outside of stressful experiences [77]. Additionally, it is unclear if a response-as-if-real measure could be used in a passive display context, as viewers are likely to be aware that any action taken would have no impact.

### 2.3.1.3 Spatial awareness

While many measures previously used for assessing VEs cannot easily be adapted to CVR due to their focus on task completion, some measures that use a post-experience task are viable. One such technique is the measurement of spatial awareness (SA) using a map placement task. In this technique, following a VE experience, participants are asked to mark the locations of objects that were visible in the VE on a map of the environment. The SA metric can be taken as the summed Euclidean distance of these objects from a ground truth. This technique has been used to compare different non-immersive representations of 360° video [78]. A related concept, spatial orientation, was used by Bowman et al. in their comparison of HMD and CAVE™ displays [79]. They found that HMD users were more likely than CAVE™ users to favour natural turning over manual turning using a joystick. Participants were therefore more likely to maintain spatial orientation in an HMD than a CAVE™ display.

### 2.3.1.4 Memory

Another objective metric that can be measured after an experience is memory. This can be achieved by asking questions following the stimulus that require the viewer to remember elements of the video. There are several different kinds of memory. Memories

can be “incidental” [80], i.e. memories made naturally during an experience, or intentional, where participants consciously try to retain memories during an experience. The popular model of memories proposed by Atkinson and Shiffrin has two memory stores: short-term memory, which lasts in the order of seconds, and long-term memory, which may be held indefinitely [81].

Measuring memory can be achieved in a number of ways, three of which are used in the Wechsler Memory Scale [82]. Free recall is a technique where participants are asked to remember elements without assistance, for example, “List as many character names as you can remember.” A second technique is cued recall, in which a related concept to the subject of the question is provided in order to aid memory retrieval. A third form, recognition, can be assessed using multiple choice questions. Testing memory immediately following an experience is an indicator of immediate recall, which relates to short and long-term memory. Evaluating retention of long-term memories requires a study that spans several days or weeks; participants must be tested after a break to measure the amount of content that has been retained.

The effect of immersion on memory is not well understood. Regan et al. showed that procedure memorisation improved with higher immersion, however their metric focused on physical tasks that relied on the improved spatial cues higher immersion provides [83]. In terms of CVR, while it might be reasonable to hypothesize that a more engaged viewer would remember more, and that a more immersive system would lead to higher engagement, this does not appear to be the case. A memory study was conducted by Rizzo et al. in their investigation of the use of 360° video for memory assessment of persons with cognitive and functional impairments [84]. Their results indicated that participants showed poorer free recall and recognition when a 360° video was watched in a HMD over a standard 16:9 TV. Explanations for this effect focused on mental load, in that participants had needed to expend mental processing to handle the complex visuals, and therefore had less available for storing memories. To ensure fairness between conditions, Rizzo et al. took all memory questions from the audio track, so participants in all display conditions received the same information. Therefore an alternative explanation for the reduced memory performance may be the novelty effect of immersive displays, as participants were distracted by the visual “wow factor” and not attending to the audio track. A similar result was found by Mania et al. in their

investigation on the impact of immersion on memory in virtual environments [85]. In their study, a 15-minute seminar was consumed in one of four conditions: audio-only, desktop, HMD, and the real world. They found that immersion level was not positively correlated with recall, however participants did remember more in the real-world condition.

#### 2.3.1.5 Narrative engagement

In their Measuring Narrative Engagement Questionnaire (MNEQ), Busselle et al. proposed a set of 12 questions that measure four aspects of engagement: narrative understanding; attentional focus; emotional engagement; and narrative presence [86]. Physiological monitoring has been used to validate this self-reported narrative engagement scale [87].

The questions in the MNEQ were distilled, through a series of experiments, from a much larger set of questions that covered an array of media engagement aspects such as empathy, narrative realism, and transportation. There is overlap between the MNEQ and questionnaires designed for interactive content, for example the Immersive Experience Questionnaire (IEQ) [88]. The MNEQ, however, is designed for passive experiences, so inappropriate aspects of the IEQ such as flow are not examined. In this context, flow is taken to mean “the state in which individuals are so involved in an activity that nothing else seems to matter” [89]. While this definition could be considered to include passive media experiences where viewers become highly engaged, the literature makes clear that flow has multiple required components, including: clear goals; direct and immediate feedback; balance between ability level and challenge; sense of personal control. Watching immersive videos is not very interactive, so arguably does not fulfil some of these aspects of flow.

An issue with a narrative engagement measure is that the stimulus may need to portray an interesting narrative story in order for the test to be valid. Due to the limited amount of 360° content available, as well as their generally short length, it may be a challenge to find immersive films which adhere to this requirement.

### 2.3.2 Evaluation of multi-view 360° experiences

In Chapter 6, a user study will be described that investigates transitions in MV360M. MV360M allows users to explore a space by actively selecting their position in the

scene from a choice of 360° views. By allowing users to navigate through a virtual space, MV360M is in some ways related to locomotion in VR. The focus of this discussion is on how drawing this parallel can help identify appropriate metrics to evaluate such systems, along with a discussion of how transitions in panoramic media have previously been evaluated in the literature.

### 2.3.2.1 Locomotion and Spatial Awareness

Exploring virtual worlds inside VR systems requires that the user be able to move within the virtual space. As there is often not a one-to-one mapping between the physical and virtual spaces, this requires techniques to be developed beyond simple physical walking. These methods collectively fall under the umbrella of VR locomotion techniques. A large number of locomotion techniques have been explored previously. Teleportation [90] is one of the most commonly used, in which a user can select a new location – for example by pointing using a hand controller – and is instantaneously moved to that new position. Another technique is auto-locomotion [91], in which the user is moved through the space based on some input, for example via a joystick. Walking-in-place is a technique in which auto-locomotion is controlled through tracked movements of the user’s body, specifically the user walking in place, and was found to create a stronger feeling of subjective presence than auto-locomotion alone [92]. Other techniques include redirected walking [93], in which physical motions are scaled to increase or decrease their effect in the virtual world, and the use of a visual metaphor such as a portal [94].

Similarly to VR locomotion, MV360M allows users to explore virtual spaces. This is a useful parallel to draw, as the VR locomotion literature has established several metrics to evaluate different techniques.

While transitions through MV360M can be framed as a VR locomotion task, there are certain differences that are important. Often, locomotion techniques are assessed on metrics such as accuracy of positioning, speed, number of collisions, and ease-of-control (e.g. as in [92]). Such metrics are not necessarily appropriate for examining different transitions. For example, the accuracy of positioning cannot reasonably be examined, as the only available locations are predetermined by the camera positions, and it is usually impossible to miss them due to the nature of the interface.

There are several metrics in the VR locomotion literature, however, that are useful for exploring the effects of transition types. Aspects such as the transition's effect on the user's spatial awareness is of importance, as content producers may wish to understand how their choice of transition will affect a user's understanding of the captured space. This may be of particular importance in MV360M, as the lack of parallax cues from head movement may have a detrimental effect on a user's spatial awareness over the six degrees of freedom generally associated with HMD experiences. Bowman et al. looked at the effect of transition types on spatial awareness in their work on viewpoint control techniques [95]. They concluded that teleportation transitions produced poorer spatial awareness than moving through the space, as assessed by the time taken to visually find a previously seen object.

Metrics from the spatial awareness literature have also been employed. Pointing tasks, in which the user is asked to indicate the direction of a previously seen object that is no longer visible, have been used in the spatial awareness literature to gauge participants' understanding of physical buildings [96] and large outdoor spaces [97]. Pointing tasks similar to these have been used in the VR locomotion literature. In work by Bowman et al., a pointing task was used to evaluate a user's ability to maintain spatial orientation while navigating through virtual corridors [98]. Their results indicated that locomotion techniques in which the user did not physically move their body still allowed them to maintain spatial orientation. Recently, Sargunam et al. used a pointing task to evaluate the effect of amplified and guided head rotations on spatial awareness [99]. Their results indicated that guided head rotations may negatively impact spatial awareness, but only found the effect to be significant for participants with significant gaming experience. As well as objective measures, subjective aspects such as the naturalness of the transition (e.g. as investigated by Usoh et al. [92]) and user preference (e.g. as investigated by Bozgeyikli et al. [90]) will have an impact on the user's experience.

As discussed by Bowman et al. in their work on VR locomotion, there is an important distinction between locomotion and navigation [95]. Navigation is a complex area that incorporates many cognitive processes. While navigation is undeniably an important concept when exploring a space in VR, like Bowman et al. we do not attempt to address the underlying processes involved, although work has previously been done

in this area [100].

### 2.3.2.2 Transitions

Transitions between panoramic images has been studied for some time. McMillan and Bishop first proposed techniques for creating novel views by interpolating between panoramic images captured from cameras with a small baseline of around 1.5m between camera locations [101]. Morvan and O’Sullivan continued this work to extend the required baseline [102]. In their work, parallax was faked using occluder masks, using a technique similar to *Tour into the Picture* [103]. Morvan and O’Sullivan also used laser scanners to create accurate models of scenes; however producing such models required expensive specialist hardware and was labour intensive, so this technique was not used in their final user evaluations. Morvan and O’Sullivan also conducted a user study to establish transition preference between faked parallax, a dip to black (fade) and a cross dissolve (blend). They concluded that blending was always preferred over fading, and that fake parallax was generally preferred over blending.

These works, however, relate to panoramic media exploration in a desktop setting. Viewing panoramic media in an immersive display such as a HMD is substantially different. For example, the user study by Morvan and O’Sullivan did not explore “cutting” (an instantaneous transition) as it was not considered to be “well suited to continuous navigation”. In an immersive context, instantaneous transitions are frequently used for navigation, largely due to the effect of vection on simulator sickness [104]. Additionally, in the work by Morvan and O’Sullivan, the orientation of the view during transitions was predetermined, which is not conducive to a HMD experience where the view is usually determined by the orientation of the HMD.

### 2.3.3 Simulator sickness

A known issue of immersive displays is the risk of simulator sickness. While the exact mechanism is not fully understood, it is believed that a mismatch between optical flow detected by the eyes and physical acceleration as detected by the inner ear can cause nausea (for a review of the literature, see [105]). In certain cases, the effects of simulator sickness can be severe. There is evidence that peripheral vision plays a central role in detecting vection, and therefore a display with a very wide field-of-view may produce a stronger effect [106]. This is of particular concern for CAVE™-like

environments, which can match the horizontal field-of-view of human vision.

Simulator sickness is also of particular concern when dealing with auto-locomotion in a HMD. This is due to the fact that simulator sickness can be induced through vection in a VR display [104], and vection is a necessary component of some forms of auto-locomotion. There have been studies that indicate that a user is less likely to experience simulator sickness if they can control or anticipate the motion [91]. Additionally, there is evidence that most users become less susceptible to particular movements with repeated exposure [107].

As a result of these factors, measuring simulator sickness is of particular importance in the experiments discussed in Chapters 5 and 6. The Simulator Sickness Questionnaire (SSQ) is the gold standard for measuring these effects, and is used extensively in the field [108].



## **Part I**

### **Media production and acquisition**



## **Chapter 3**

# **Media acquisition**

In order to perform studies on the evaluation of 360° media, it was first necessary to obtain 360° media content with which to perform these experiments. As the landscape of the 360° media industry has changed substantially over the last four years, this may now appear to be an almost trivial issue for CVR content. Four years ago, however, high-quality 360° videos were not readily available. As a result, a focus at the beginning of the project was on creating high-quality content that could be used in future experiments. In this chapter, we briefly discuss this process, performed in collaboration with BBC R&D.

In the latter stages of this project, high-quality 360° content became readily available online. As a result, our focus shifted from creating 360° media to selecting appropriate content from the large amount available. YouTube began to support 360° video in 2015 [109]. This, coupled with an explosion of interest in the medium, led to the availability of a plethora of content. In the second half of this chapter, we discuss the selection process used to determine which videos would be most useful for evaluating immersive viewing experiences. The complexities of producing MV360M are also discussed, as well as the decision to use third-party content for the study performed in this area.

## **3.1 BBC in collaboration with UCL**

### **3.1.1 Test documentary**

A short test documentary was filmed by the BBC in collaboration with UCL. The video capture hardware, operation and stitching was provided by UCL, while audio capture was provided by the BBC. The writer and director was Peter Boyd Maclean. Maclean



**Figure 3.1:** Filming with the Point Grey Ladybug3 on the roof of OnEustonSquare

has worked extensively with Nonny de la Peña, who is widely recognised for her work in immersive VR journalism. An image of the camera being set up can be seen in Figure 3.1.

#### 3.1.1.1 Equipment

The filming was conducting using UCL's Point Grey LadyBug3 camera in February 2015. The Ladybug3 uses 6 cameras with wide angle lenses to capture panoramic content. It comes complete with stitching software that can output 360° images or video. The lack of a camera facing downwards means that the base of the sphere is not captured – this is seen as a black hole below the camera.

To capture content outdoors the Ladybug3 was connected to a 17-inch MacBook Pro with a FireWire 800 port, as shown in Figure 3.2. The Ladybug software is Windows only, and VMWare Fusion cannot virtualise FireWire ports, so the laptop was booted into Windows. The MacBook itself was capable of supplying enough power to the camera even when not plugged in to the mains. Without an external power supply, approximately 30 minutes of filming could be performed on a single charge.

The Ladybug3 was set to 15 frame per second (fps) using JPEG 12-bit compression and full resolution of 5400x2700 pixels. The Ladybug3 is capable of capturing at up to 32fps, but only if half vertical resolution is used. At full resolution, 16fps is the maximum frame rate.



**Figure 3.2:** Point Grey’s Ladybug3 connected to 17-inch MacBook Pro via Firewire 800

### 3.1.1.2 Software

Following advice from the BBC’s Richard Taylor, stitching was performed using the Point Grey software into a sequence of PNG images. These images were then turned into a video file using ffmpeg. At high resolution and using a good colour mode, the stitching process can be quite lengthy. Stitching was performed on a Mid-2010, 17-inch MacBook Pro with an Intel Core i7 running at 2.66GHz, with 8GB RAM and using a TS256B Apple SSD hard drive. Stitching on this device using the “High Quality Linear” colour interpolation mode, one minute of video took around two hours to stitch at a resolution of 5400x2700.

When synchronising the audio with the output from the Ladybug, it became apparent that the frame rate was inconsistent, varying across the duration of the video. As ffmpeg assumes a constant frame rate, this resulted in erratic playback speed. Settings can be chosen that reduce the Ladybug’s data production rate, and the MacBook had an internal SSD. Despite this, occasionally frames were still dropped. When syncing audio even a single dropped frame was noticeable.

It was possible to test for dropped frames in a recording using the LadybugCapPro software provided by Point Grey. This provided a report of how many frames were dropped and where. This information was used to correct for the missing frames. A Ruby script was created that parsed out the missing frame details from the report produced by the LadybugCapPro software, and used that data to copy neighbouring frames to fill the gaps. While this works, it adds an additional step to the post-production process.



**Figure 3.3:** Frame from test documentary shot with Point Grey’s Ladybug3 camera.

### 3.1.1.3 Results

Filming with the Ladybug3 presented many challenges. The short filming time provided by a single charge presented practical issues for filming outdoors. The tethered laptop presented issues of form factor. The dropped frames required additional work in post production to ensure a consistent frame rate. Additionally, the low frame rate of 15 frames per second would be considered unacceptable for most viewers and is dramatically below the current standard for video production. A frame from the documentary can be seen in Figure 3.3. Frames from this video were used in the investigation of object removal in panoramic media, discussed in Chapter 4.

## 3.1.2 The Resistance of Honey

The next piece BBC R&D worked on was “The Resistance of Honey”, a 360° documentary video about a beekeeper filmed in September 2015. A frame from this video is shown in Figure 3.4. Due to the limitations of filming with the Ladybug3 described above, it was decided to use different camera hardware. An array of GoPro HERO4 cameras held using a Freedom360 mount, similar to that shown in Figure 3.5, was made available by Middlesex University. Sound equipment was provided by the BBC. As a result of Middlesex University’s involvement, UCL played a smaller role in this production. The piece was written and directed by Peter Boyd Maclean, with UCL performing equipment preparation and on-set operation. Stitching and editing was done by Dr Peter Passmore of Middlesex University, and Peter Boyd Maclean. This piece would go on to be the first 360° video by BBC R&D that was publicly released. An



**Figure 3.4:** Frame from *The Resistance of Honey*.



**Figure 3.5:** GoPro array using a Freedom360 mount.

edited version of this video was used in the study described in Chapter 5.

## 3.2 BBC and external companies

Many other pieces of 360° footage filmed by the BBC and external companies were made available for use in the work undertaken in this thesis. This included several test shots by the film maker Peter Boyd Maclean, which featured outdoor scenes such as parks, and a further short documentary shot by the BBC. The BBC also made available some high-quality footage shot in Broadcasting House, which included footage from the BBC newsroom. Frames from some of these video files can be seen in Figure 3.6. Frames and video from this content were used in the investigation of object removal in panoramic media, discussed in Chapter 4.



(a) Frame of a London park. Video courtesy of Peter Boyd Maclean



(b) Frame of Westminster. Video courtesy of BBC R&D



(c) Frame from BBC newsroom. Video courtesy of BBC R&D

**Figure 3.6:** Examples of 360° video made available by the BBC and Peter Boyd Maclean.

### 3.3 YouTube and the growth of 360° video

YouTube added the ability to share 360° content via its platform in March 2015 [109], with Facebook following suit a few months later [110]. Improvements in the production pipeline, as well as growing interest in the medium, meant that around this time an enormous amount of high-quality, diverse content was becoming available. As a result of this, content creation was no longer an essential component of experience evaluation.

Additional CVR content was found, via YouTube, that conformed to the requirements necessary to measure various aspects of the end-user experience. This included aspects such as how frequently the video cut between different locations, character placements, and the strength of the narrative arc. These requirements are discussed fully in Chapter 5. A complete list of all content used can be found in Appendix B.

### 3.4 Multi-view 360° media

In Chapter 6, a user study is discussed in which MV360M was investigated. This required the use of MV360M, which is currently much less common and harder to access than single-view 360° media. Additionally, as will be discussed further in Chapter 6, it was important for the study that a 3D model be available of the scenes depicted in the media.

At first it was attempted to create geometry proxies from existing MV360M. BBC R&D provided a MV360M piece they had recently captured. Manually creating geometry proxies using this media, however, proved to be extremely labour intensive, with unsatisfactory results. Automatic camera calibration seemed unlikely to work, due to the extremely wide baselines involved, and the fact that the equirectangular images in question were in fact multiple images blended together, each with its own extrinsic and intrinsic camera matrices. Recent tooling improvements in interactive geometry proxy construction from panoramas may make this process easier in the future, however these systems are currently limited to single-view 360° media [111].

As a result of these issues, it was decided to approach a third-party source to obtain MV360M for which 3D models were already available. Matterport, a California-based property technology company, have created a camera that captures 360° RGB-D images. This camera, shown in Figure 3.7, can be used to scan indoor scenes, taking 4K panoramic images from specific locations, while also building a 3D model of the



**Figure 3.7:** Matterport's 360° RGB-D camera



**Figure 3.8:** An example 3D model from Matterport

scene. This camera is only capable of capturing static scenes. Matterport were able to supply us with data for three scenes, an example of which is shown in Figure 3.8. These scenes were used in the MV360M user study, discussed in Chapter 6.

### 3.5 Conclusion

In this chapter, the process by which 360° media was acquired for all stages of the project was discussed. Initial experiments in capture using the Ladybug3 camera were covered, and issues connected with this technology highlighted. Collaborative effects

between BBC R&D, Middlesex University and UCL to create the publicly released “The Resistance of Honey” documentary were also detailed. Additionally, the resourcing of single-view 360° media from YouTube and multi-view 360° media from Matterport were also described. A complete list of all 360° media used throughout this work is available in Appendix B.

## Chapter 4

# Object removal in panoramic media

A project was undertaken that explored object removal in 360° images. This chapter describes that project, and is reproduced from sections 4–7 of a paper that was published in the 2015 *Proceedings of the European Conference on Visual Media Production* (CVMP) [16]. First, a method of object removal in 360° images is described, in which field-of-view expansion is combined with Graphcut Textures to remove objects. Secondly, inpainting in 360° images is examined. Finally, the latter technique is shown to work for video in certain situations.

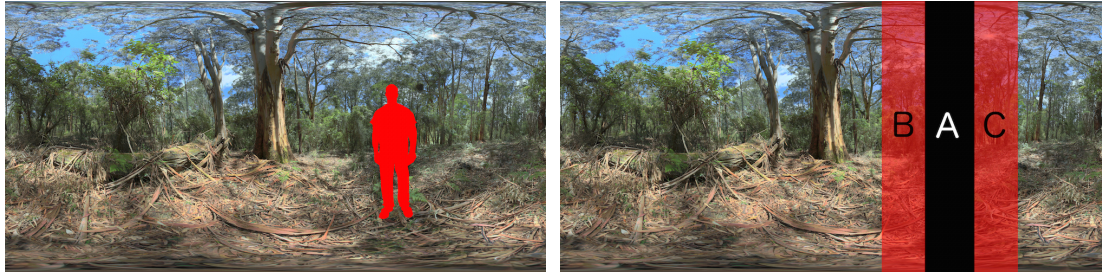
## 4.1 FOV expansion and graph cuts

In many cases, object removal in a 360° panorama can be formulated as FOV expansion. By expanding the HFOV beyond 360°, an area of overlap is introduced to the viewing sphere. If the overlap is positioned at the point of the unwanted object, a cut between these ends can be done in such a way as to remove the object.

### 4.1.1 Algorithm formulation

This algorithm can be formulated using an equirectangular projection [31]. This projection has the useful property that increasing the HFOV can be achieved by simply increasing the image width.

An example of this method is shown in Figure 4.1. The section of the equirectangular image containing the object is removed. As equirectangular images have a distortion pattern that is constant horizontally but varies vertically, the removal is performed vertically to ensure content maintains the correct variation of vertical distortion. At this stage, the image could be stretched to restore the 2:1 aspect ratio and considered



(a) Equirectangular with object to be removed in red. (b) Section A is removed, while sections B and C are overlapped and a good transition found using Graphcut Textures.

“Sherbrooke Forest” by Peter Gawthrop, used under CC BY-NC 2.0

**Figure 4.1:** FOV expansion and Graphcut Textures

a 360° panoramic image. However, the introduced cut would be highly noticeable and jarring. To disguise the cut, Graphcut Textures can be used [33].

To create a plausible join, the left hand side and right hand side of the join are overlapped as shown in Figure 4.1b. A larger overlapping region provides more scope for a good cut to be found, however it also removes more of the original panorama. This is a parameter that can be altered depending on the context. Graphcut Textures is then used to find a good join between these overlapping sections, formulating the problem as a min-cut/max-flow optimisation that favours cuts at areas of similarity.

### 4.1.2 Results

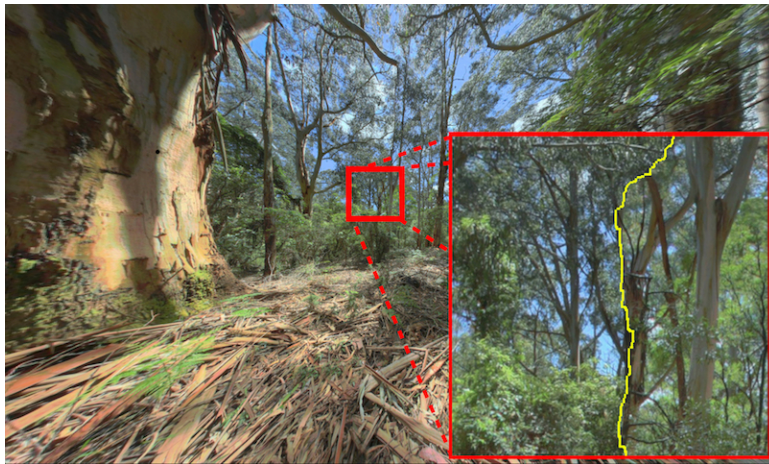
We tested the above technique on equirectangular images. As can be seen in Figure 4.1a, we digitally added an object to the scene that presented a substantial challenge to remove. We performed the technique as described in section 4.1.1, the results of which can be seen in Figure 4.2.

The results are fairly compelling. The offending object is completely removed, and the cut is well disguised as can be seen in Figure 4.2c. Some artifacts can be seen, for example the branch at the centre top of image 4.2c ends abruptly where the cut has taken place. We have also needed to remove more of the original panorama than in the copped version seen in Figure 4.2b to allow an overlapping region for graph cuts to take place.

Two further examples can be seen in Figure 4.3. Here, frames from two videos are processed to remove crew and equipment. The removal in Figure 4.3c is challenging as the object is fairly wide in the scene, and the background change is nontrivial. To



(a) Rectilinear view with object to be removed in red. (b) A simple straight cut: the cut appears as a noticeable sharp edge.



(c) The proposed graph cuts method: using Graphcut Textures, the cut is well disguised. In the closeup the cut is shown in yellow.

“Sherbrooke Forest” by Peter Gawthrop, used under CC BY-NC 2.0

**Figure 4.2:** Rectilinear views of image 4.1a following HFOV expansion and region removal. As a large FOV has been used to display more of the scene, stretching distortion can be seen at edges.

produce this cut, the section marked to be removed included the crew and most of the bag – the remainder of the bag was removed automatically during the graph cut phase. Importantly, these cuts took place in non-salient areas of 360° panoramas. This means that the viewer is most likely not focusing their attending in this direction, and therefore some minor artifacts may go unnoticed.

### 4.1.3 Limitations

By removing a section of the panorama and stretching the remaining content to restore the 2:1 aspect ratio, a circular distortion centred around the poles has been introduced. This circular distortion is noticeable even in natural scenes, as shown in Figure 4.4. The amount of the panorama removed will affect the results, with smaller cuts introducing



(a) Before cut: the director is visible and should be removed.



(b) After cut: the seam is not obvious, but minor artifacts in shadows.

Frame courtesy of Peter Boyd Maclean



(c) Before cut: some crew and equipment are visible.



(d) After cut: minor artifacts are noticeable in the sky

Frame courtesy of the BBC

**Figure 4.3:** Rectilinear views of equirectangular video frames, showing object removal using the proposed graph cuts method.

less distortion. The distortion effect in Figure 4.4 is quite pronounced as 25% of the HFOV of the original panorama has been removed.

This distortion effect alters the content in such a way that straight lines are no longer guaranteed to be straight. This is particularly pronounced near the poles. This effect may or may not be noticeable depending on the content; natural scenes such as that of Figure 4.1a may appear plausible, while artificial structures with straight lines will make this effect more obvious, as shown in Figure 4.5. As we will discuss further in section 4.2.1, this effect can be reduced by scaling the content in a non-homogeneous fashion.



(a) Before cut: the south pole appears normal.

(b) After cut: a circular distortion has been introduced, centred on the pole.

“Sherbrooke Forest” by Peter Gawthrop, used under CC BY-NC 2.0

**Figure 4.4:** Rectilinear views of the circular distortion introduced by the proposed graph cuts method at the south pole.



(a) Before cut: straight lines near the north pole appear straight.

(b) After cut: straight lines near the north pole appear curved.

**Figure 4.5:** Rectilinear views of the tops of buildings, showing distortion introduced at the north pole by the proposed graph cuts method. Note that the removed section is behind the viewer, so cannot be seen in these images.

Marking the region to be removed is very easy. Only the x-axis start and end positions of the region to be removed, as well as the overlap size, need to be specified. This makes the process quick compared to the creation of even a rough mask. However, this speed comes at the expense of control. The entire marked region will be removed and the overlap searched for a good cut – the user cannot specify regions in the overlapping area they would like to keep. As the system takes only a few seconds to create the new panorama, a trial and error approach can be used to identify good settings for an image interactively.

The graph cut method used suffers the same limitations as Graphcut Textures.



(a) Before cut: the person in the red coat should be removed.



(b) After cut: background structures are noticeably incorrect.



(c) Before cut: the teleprompter should be removed.



(d) After cut: a good cut cannot be found.

Frame courtesy of the BBC

**Figure 4.6:** Failure cases of the graph cuts method: rectilinear views before and after cuts. The objects are successfully removed, however surrounding structures are incorrect resulting in unrealistic images.

Sometimes no decent cut is available, and the two sides cannot be realistically combined. This can happen when a significant change happens in the background across the object to be removed. Additionally, artificial structures and buildings are difficult to cut plausibly as they tend to have a regular shape with many straight lines, which are challenging to join realistically. An example of this can be seen in Figure 4.6. For this reason outdoor, natural scenes are likely to produce better results.

A cut alters the physical layout of the space. In situations where the viewer does not have a strong mental model of this layout, this is usually not an issue. In some cases, however, such as a square room that is no longer square following the cut, the alteration is quite pronounced and slightly disconcerting. In scenes with very regular surrounding structure, the technique of section 4.3 may well be applicable.

## 4.2 Extensions & refinements

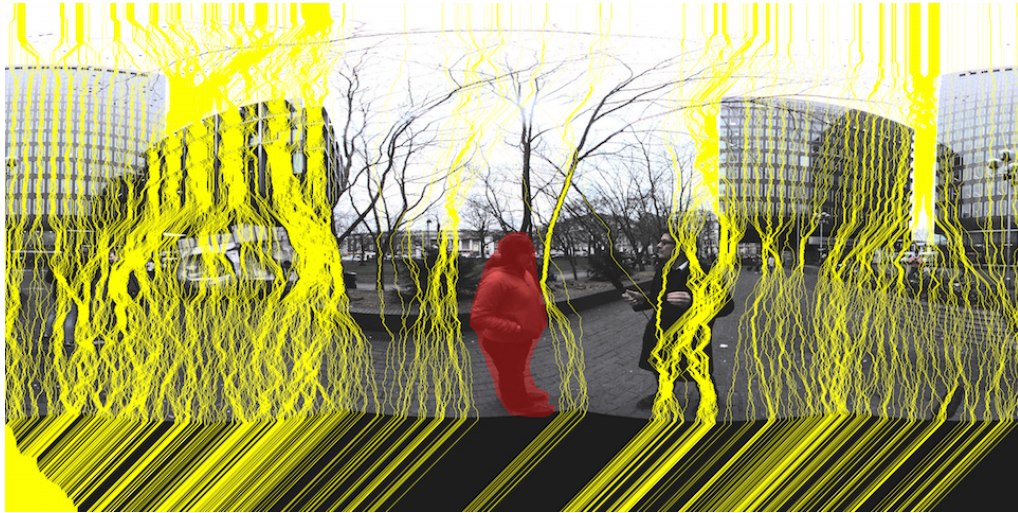
### 4.2.1 Retargeting techniques

The method described in section 4.1 performs homogeneous stretching of content to restore the 2:1 aspect ratio of the equirectangular image following a cut. While this can produce good results, it can also cause noticeable distortion. For example, following the removal of a large object, people are visibly more stretched in the horizontal axis, and buildings with straight lines close to the poles no longer appear straight. Retargeting techniques can be used to reduce these distortions.

Seam-carving can be applied to prevent the distortion of important sections. These sections, for example people or straight lines near the poles, could be detected automatically in panoramic images using the methods proposed by Sacht et al. [45]. However for simplicity, the example presented here was produced using a manually created saliency mask. As shown in Figure 4.7, additive seam carving was used to increase the width of an image while preserving the salient element, in this case a person who is expected to be the focal point.

Seam carving in panoramic content may be even more effective than in standard format media due to the large amount of non-salient areas. Standard media has a comparatively small FOV, so often the entire screen is filled with salient content. Panoramic media, in contrast, generally has a large amount of non-salient content. However, seam carving is a computationally expensive process in this context due to the extremely large size of panoramic images. Carving 600 seams in the moderately sized 2400x1500 frame shown in Figure 4.7a took 15 minutes, despite the use of a well known, fast implementation [36]. Additionally, seam carving can introduce noticeable distortions in some scenes, particularly regular structures and straight lines. Instead, a simpler method can be used.

By performing stretching of the non-salient content only, good results can be



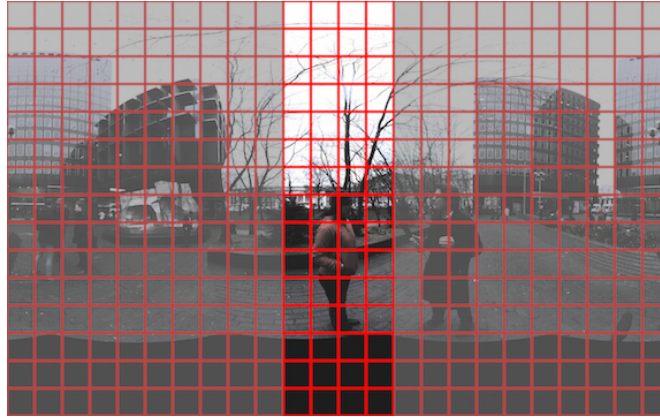
(a) Equirectangular undergoing seam carving to restore 2:1 aspect ratio following large object removal. Saliency mask is highlighted in red, carved seams are shown in yellow.



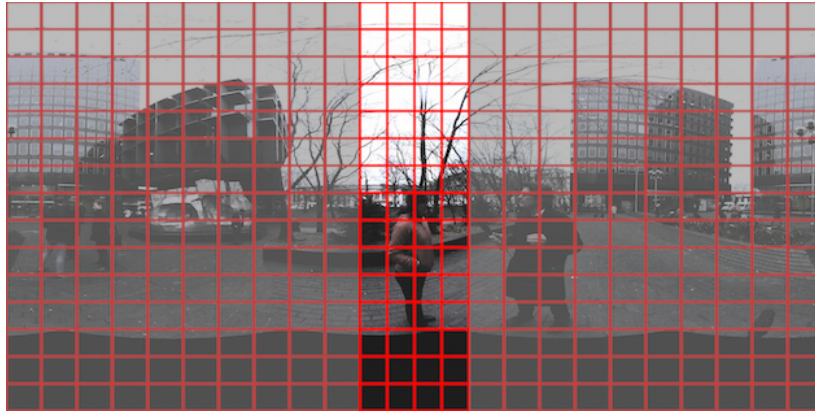
(b) Rectilinear view: before the cut. (c) After cut: homogeneous image stretch. (d) Seam carved: person preserved by mask.

**Figure 4.7:** Seam carving can be used to preserve salient content during FOV expansion. Note that the removed section is behind the viewer, so cannot be seen in images b–d.

achieved at interactive speeds. After manually determining the salient areas, non-salient content is stretched horizontally to restore the 2:1 aspect ratio following a cut, as can be seen in Figure 4.8. As we are only stretching along the x-axis and a large amount of non-salient content is available to perform the operation easily, a very simply method can be used. The sections to the left and right of the salient area are stretched, while the central region containing the salient content retains its original size. If a more complex scaling was required, for example in scenes with many disconnected salient areas, the work of Wolf et al. may facilitate a more appropriate non-homogeneous stretch [38]. Similar results to those of Figure 4.7 are achieved, in less time and with less introduced distortions in areas with regular structure.



(a) Equirectangular following large cut.



(b) Equirectangular is non-homogeneously stretched in the x axis to restore 2:1 aspect ratio. Middle area is not stretched to preserve central figure.

**Figure 4.8:** Grid shows non-homogeneous stretching used to preserve salient content. In these images, the stretched sections have been darkened to highlight areas that were altered.

### 4.2.2 Tripod removal

The method outlined in section 4.1 can remove objects near the equator of the image. To remove objects above or below the camera, such as the tripod, an adaptation must be made. Removing the tripod is achieved by rotating the panorama so the object appears near the equator, removing the object as in section 4.1, and then reversing the rotation to return the panorama to its original orientation. Rotating the panorama can be thought of as choosing different poles on the viewing sphere before projecting to equirectangular, as shown in Figures 4.9a and 4.9b. However, some issues are introduced by this method of performing cuts.

As indicated in section 4.1.3, a circular distortion is introduced at the poles during the cut. If the poles during the cut are in fact from the equator of the original panorama, these distortions become highly noticeable. As can be seen in Figure 4.9e and 4.9f, the

horizon becomes bent. There are a number of ways to mitigate this effect. The first is to remove an equally sized cut from the opposite side of the sphere. This means the horizon stays in place. This can produce plausible results in some cases. However in some cases this will only increase the perceived distortion, particularly when there are straight lines such as buildings, such as in Figure 4.9g.

An alternative approach is to stretch the content in a non-homogeneous way as described in section 4.2.1, specifically, stretching only the content below the horizon to return the 2:1 aspect ratio following a cut. This allows the horizon to remain flat, and preserves the content above the equator at the expense of some additional stretching below the equator. An example of this can be seen in Figure 4.9h.

It is important to note that this method of tripod removal requires two rotations of the sphere to be stored as equirectangular images – the first to move the pole to the equator, and the second to move it back after editing. Each of these rotations uses a filter that distorts content at the poles, resulting in reduced clarity of the image. A way to mitigate this is to work in a higher resolution than required for the final image before scaling down, although this will also increase computation time.

## 4.3 Inpainting

The method described in section 4.1 changes the FOV of the image, resulting in warping of the content. Additionally, to perform tripod removal, the poles must be moved twice, each using a filter that increases blurring. Instead, inpainting could be used to remove unwanted objects or fill holes. To perform inpainting, the object or hole to be removed is masked to identify the area to be inpainted. Content from the rest of the image is then used to fill this area in a way that appears plausible. In all of the examples in this document, the mask was created manually, and the inpainting performed using Adobe Photoshop’s content-aware fill [43].

### 4.3.1 Inpainting in equirectangular

The most simple method to inpaint objects in panoramic images is to inpaint directly in equirectangular. In many cases, this will produce excellent results and no further work is required. Examples of this can be seen in Figure 4.10. However there are many situations in which this method will not work. Inpainting the tripod, for example,



(a) Original equirectangular – tripod is stretched across the south pole.



(b) The sphere is rotated so the tripod is at the equator.



(c) Before the cut: looking down, the tripod is visible.



(d) After cut: tripod removed using FOV expansion and graph cuts.



(e) Before the cut: the horizon is straight.



(f) After cut: circular distortion at the pole has warped the horizon.



(g) Straightening the horizon: taking an equal sized cut from the opposite side of the sphere.



(h) Straightening the horizon: non-homogeneous stretching of content preserves buildings better.

Frame courtesy of the BBC

**Figure 4.9:** Tripod removal using FOV expansion and graph cuts.



(a) Closeup of object to be removed. (b) Inpainted in equirectangular.

Frame courtesy of Peter Boyd Maclean



(c) Closeup of object to be removed. (d) Inpainted in equirectangular.

Frame courtesy of the BBC

**Figure 4.10:** Inpainting directly in equirectangular.

is infeasible using this method as the tripod is normally at the south pole – an area so distorted in an equirectangular projection that inpainting cannot produce plausible results. Indeed, even masking the tripod for inpainting is a challenge as the distortion is so strong that it is difficult to identify it definitely.

One method to combat these issues is to rotate the tripod to the centre of the equirectangular image, perform inpainting, and then rotate the panorama back to its original orientation. These rotations can be performed in a similar fashion to section 4.2.2. In this case, however, the effects of the filter used during rotation can be undone at the end of the process. This is achieved by rotating the mask created during the inpainting step through an identical rotation as that of the image. This mask is then used to copy the inpainted material from the final image into the original, unrotated image. To prevent noticeable discontinuities the copied section can be feathered in. As can be seen in Figure 4.11, this method is capable of producing excellent results.

An issue with inpainting in an equirectangular image is that it can produce poor

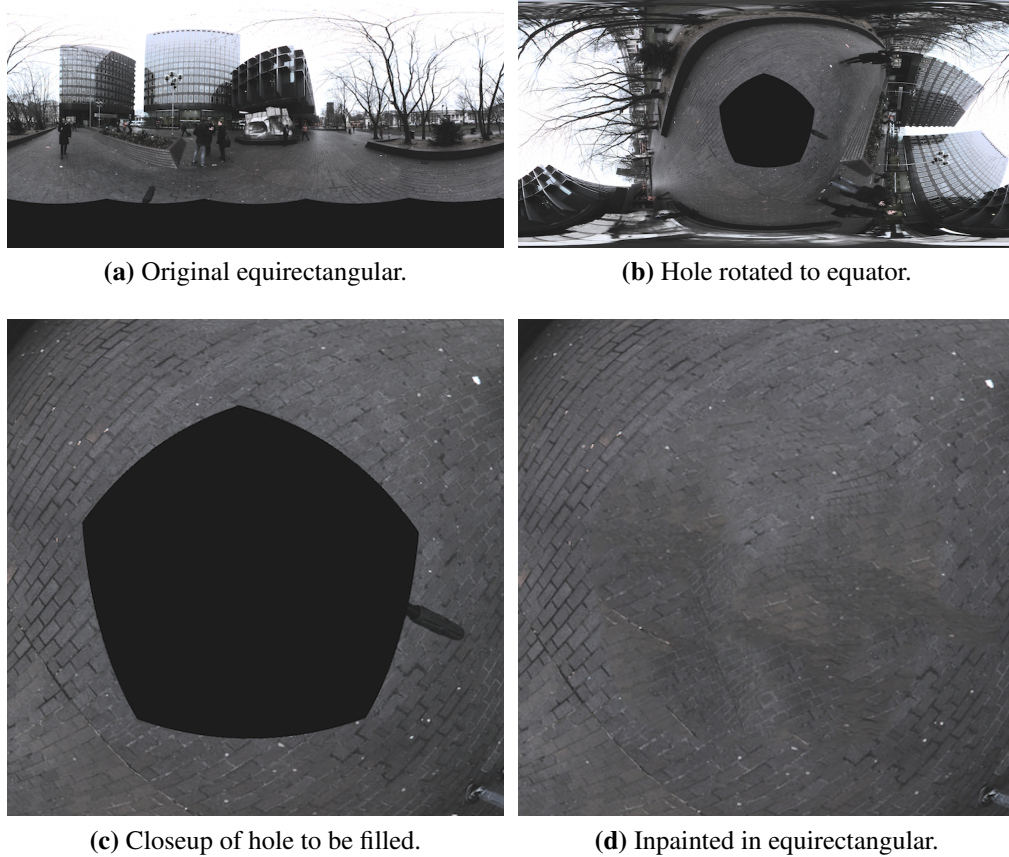


**Figure 4.11:** Inpainting in equirectangular following rotation of south pole to the equator, as in Figures 4.9a and 4.9b.

results in some situations. Content can be used for inpainting that does not match the distortion of the hole being filled. This is particularly noticeable when inpainting geometric textures. To highlight this, a challenging issue was considered. Some 360° cameras do not capture a full sphere. Point Grey’s Ladybug3, for example, does not have a downward facing camera. This results in a black hole covering the south pole. Removing this black hole follows a similar process as removing an unwanted object – a section of the media must be removed or replaced while maintaining a plausible visual result. As can be seen in Figure 4.12, inpainting this large hole with geometric content causes issues. In Figure 4.12d, brickwork from elsewhere in the equirectangular image is used, resulting in an image that is noticeably incorrect.

### 4.3.2 Straight line preserving projections

A possible solution to the problems seen in Figure 4.12d is to inpaint in a projection that preserves straight lines. A cubic projection is one in which the sphere is projected onto the six sides of a cube. Each cube face is a rectilinear image with a HFOV and VFOV of 90°. While this projection has issues in terms of storage and display – being somewhat complicated to understand when viewed in flat form – it has the advantage of having little distortion on any of the cube’s faces. Importantly, straight lines appear straight in each tile. By inpainting using only the bottom tile of a cubic projection, the hole shown in Figure 4.13a can be removed with fairly good results, as shown in Figure 4.13b. However, using only the bottom tile of the cube means that less of the panorama is available for matches to be located. For many cases this would not be an issue as the

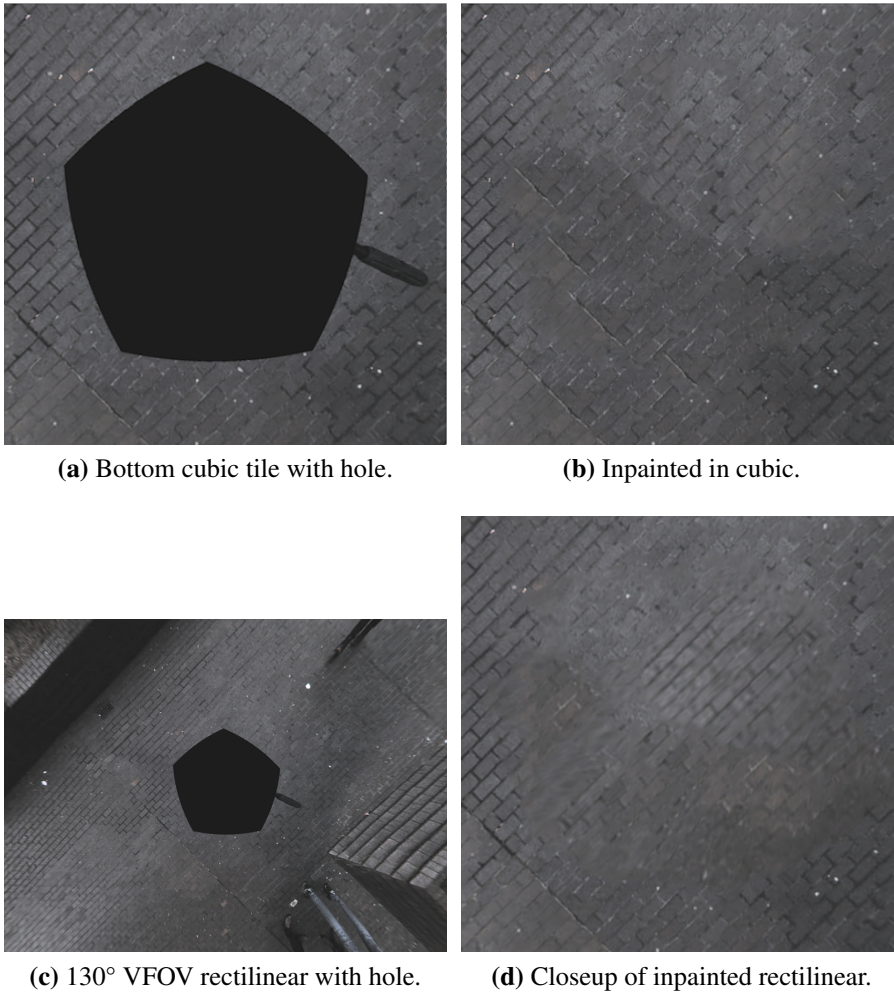


**Figure 4.12:** Failure case of inpainting in equirectangular: large hole at south pole inpainted with geometric texture.

best content to inpaint with will likely be near the hole. For complex inpainting tasks requiring more information, another approach may improve results.

The inpainting of geometrically complicated surfaces is an example of a case where improvements may be possible over a cubic projection. This can be seen in Figure 4.13d, where the  $130^\circ$  VFOV rectilinear image centred on the hole shown in Figure 4.13c is inpainted. At  $130^\circ$  VFOV, content at the edges of the rectilinear image undergo substantial stretching distortion. In the cubic version in Figure 4.13b, the results are crisp but there is noticeable repetition. In Figure 4.13d, repetitions are less obvious, although this comes at the expense of clarity as the results are more blurred. This makes sense – the rectilinear version in Figure 4.13c has more choice as it covers a larger area, however stretching distortions at the edges of the rectilinear mean blurred content is available for use by the inpainting algorithm.

Whether or not a cubic projection is sufficient will depend on the size of the hole and the type of content being inpainted. Inpainting could be performed on rectilin-



**Figure 4.13:** Inpainting of large hole at south pole.

ear views with differing FOVs, allowing the best result to be found. Inpainting the 90° HFOV/VFOV rectilinear tile of the cubic projection has the advantage that the inpainted tile can be swapped in for the original tile and no more work is required. For rectilinear views with other FOVs, the inpainted rectilinear content must be copied back into the original panorama, adding an additional step to the process. Working in a higher resolution than required for the final panorama may also be useful, as stretched content at the edges of the rectilinear will be better quality and therefore produce superior results if used by the inpainting algorithm.

### 4.3.3 Limitations

Inpainting panoramic content can produce good results in many cases. It suffers similar limitations, however, to inpainting standard format content. There are times when inpainting cannot produce a plausible result. Examples of this can be seen in Figure 4.14.

In Figure 4.14b, the implausible removal of a tree's trunk mean the resultant image is unconvincing. In Figure 4.14d, revealed structure cannot be realistically constructed using content found elsewhere in the image.

In such cases, it may be better to remove the area entirely using the method described in section 4.1, as seen in Figure 4.14e. In some cases, however, neither method will work, as shown in Figure 4.14f where a camera operator cannot be removed while retaining a plausible image. The substantial change in background and the geometric structure of the surrounding room prevent the successful application of the graph cuts method.

The completion of the nadir using inpainting shown in Figure 4.13 is very successful, even in the presence of complex geometric textures. However, this is in part due to the fact that the content is viewed fronto-parallel, as the camera's downward vector is perpendicular to the ground plane. To inpaint geometrically textured content on surfaces that are not viewed fronto-parallel, it may be beneficial to correct the perspective distortion prior to inpainting, as in the work of Pavić et al. [112].

## 4.4 Video

Extending the technique described in section 4.3 to video is easy in certain situations. If the camera is fixed, the background is static, and foreground objects do not pass in front of the inpainted hole, the results of inpainting in one frame can be copied to the other frames. Small changes in global illumination can be handled by adjusting the brightness of the inpainted section to match the target frame before transfer.

While these constraints may seem restrictive, tripod removal for a static camera often fulfils the required criteria. This assumes other foreground objects or shadows do not cross the inpainted section. This method was used to remove the tripod from a video sequence, the results of which can be seen in Figure 4.15.

The object removal technique described in section 4.1 could be extended to video in a similar way, applying the cut found in the first frame to each successive frame in the video. This method would suffer from similar limitations as well – foreground objects could not move across the cut area, and dynamic backgrounds would not be supported.



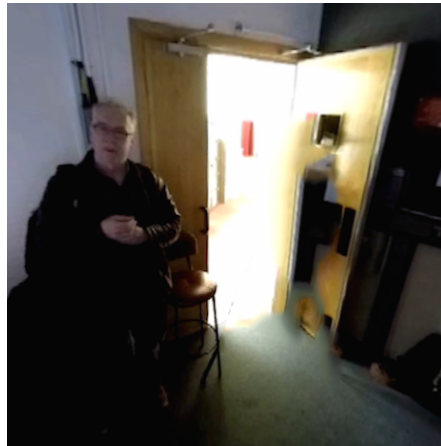
(a) Closeup of large object to be removed.



(b) Revealed tree trunk is not reconstructed by inpainting.



(c) Camera operator should be removed.

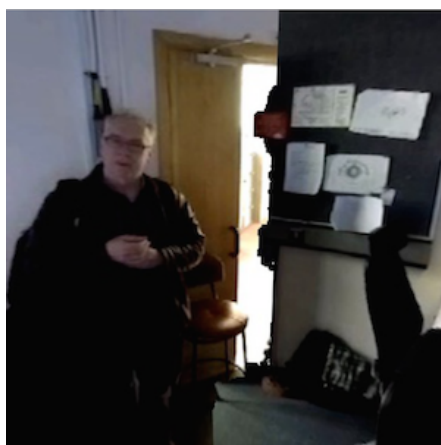


(d) Inpainting does not realistically complete the door.

Frame courtesy of the BBC

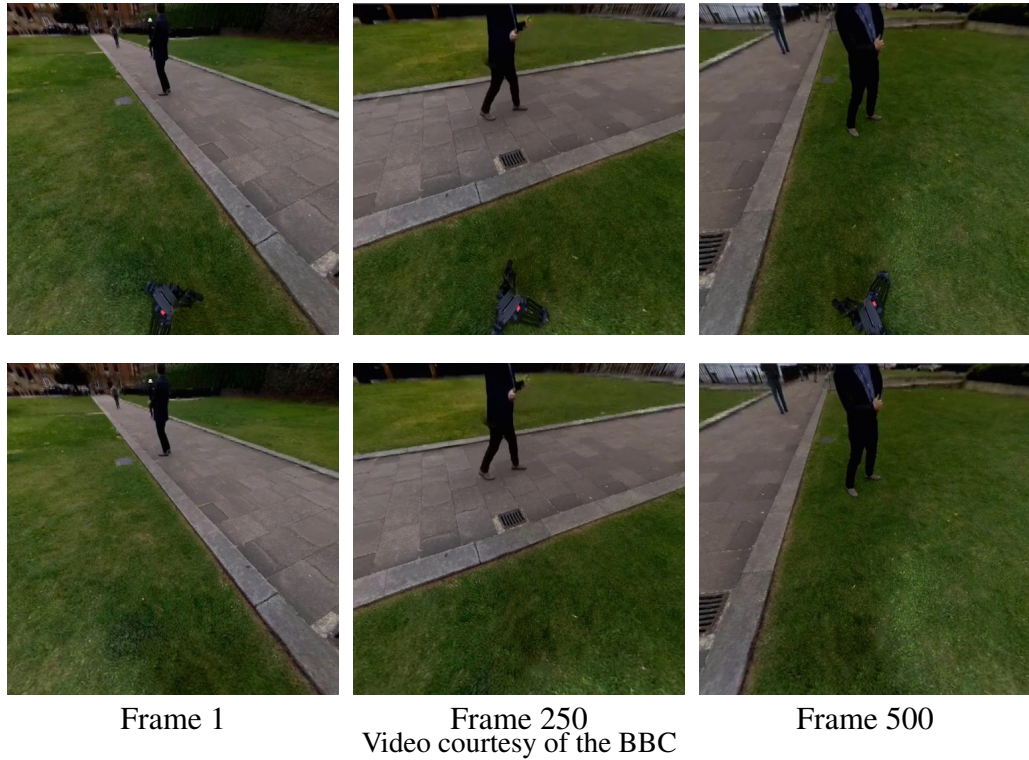


(e) Graph cuts applied to image 4.14a: the result is better than 4.14b.



(f) Graph cut applied to image 4.14c: the result is worse than 4.14d.

**Figure 4.14:** Failure cases of inpainting method. Inpainting does not always produce plausible results when the removed object is very large or novel shapes are revealed. The proposed graph cuts method can improve results in some cases.



**Figure 4.15:** 100° FOV rectilinear views of an in-painted video. The in-painted section was copied between the equirectangular frames. Above, the tripod can be seen at the bottom of the unedited frames. Below, the tripod has been successfully removed.

## 4.5 Conclusion

It has been shown that two methods for object removal are viable for use in 360° panoramas. Inpainting can produce excellent results in many situations. In many cases where the object to be removed is not at the poles, it was shown that the standard equirectangular projection can be used without additional steps. Due to the distortion characteristics of equirectangular images, however, more needs to be done to inpaint content at the poles. Tripod removal was shown to work well by adding rotation steps to the process, while cubic and rectilinear projections are required for complex inpainting tasks such as geometric textures. Inpainting failed to produce good results, however, in some cases where the background being revealed could not be plausibly reconstructed using content in the image.

In these cases, the FOV expansion and graph cuts method described in section 4.1 could be used. Instead of attempting to reconstruct the background revealed during object removal, the entire section is removed and the neighbouring content cut together in a plausible way. This method was shown to produce good results including when the

object being removed was not at the equator. Methods were also discussed to alleviate the distortions this method can introduce, such as seam carving and non-homogeneous warping. However, in some situations such as the presence of very regular surrounding structure, this method may fail to produce acceptable results.

## **Part II**

### **360° media evaluation**



## Chapter 5

# User study: evaluating the effect of display type on the viewing experience for panoramic video

This chapter looks at monoscopic, passive, fixed-viewpoint 360° videos, as these are by far the most commonly available type of video for virtual reality. These videos span a broad spectrum of genres, from news and journalism to comedy and horror.

There are several issues with HMD-based playback of 360° videos. Arguably the most detrimental problem is the lack of directorial control over what the viewer sees, as the director cannot dictate in which direction the viewer is looking at any given time. This presents issues for narrative understanding, and may lead to the viewer feeling they have missed important elements. In a 2016 user study, “audiences expressed FOMO (what the kids are calling ‘fear of missing out’)” [113]. Additional issues include the physical discomfort of wearing the headsets, and the feelings of vulnerability and unsociability generated by being cut off from the physical world.

The SurroundVideo+ (SV+) display was designed to mitigate some of the issues associated with HMD playback, while attempting to retain some of the immersive characteristics. SV+ is an immersive display, featuring a TV as a focal point to provide a directed viewing experience, as well as projection-based peripheral content to provide immersive visuals. The peripheral projections in our SV+ are provided by a CAVE™-like display [114]. The use of projection can allow SV+ to provide a social viewing experience. The use of projection also means viewers do not need to wear any equipment, and that no setup is required following initial calibration. By using projection

mapping and radiometric compensation techniques, SV+-like systems could be one model for the living room of the future. The use of a CAVE™-like display can be seen as an ideal version of SV+ where any furniture or fixings in the room are effectively imperceptible or irrelevant in the context of the video presentation.

As well as presenting issues, immersive displays may provide new opportunities. HMDs provide a new way to view media that may provide novelty and additional enjoyment. For example, it has been suggested that “VR horror experiences can be much more intense, isolating and terrifying than when played on a standard 2D display” [115]. While this statement was directed at VR gaming, CVR is likely to be similar. The complete occlusion of the real world in HMD experiences gives media producers the opportunity to control many elements of the experience, allowing them to create highly atmospheric content. Additionally, the isolation and physical vulnerability of viewers would likely increase any sense of fear. While this will likely produce more powerful horrors, it has been suggested that “jump scares” in VR may be too intense, prompting Oculus VR to “strongly discourage” content creators from using them [115].

Evaluation of passive CVR experiences presents several issues. Task completion is almost always used in HCI evaluation, with metrics such as speed or accuracy forming the basis of most analyses. As CVR does not involve active tasks, these frameworks are not suitable. Evaluation techniques used in media psychology, however, do not specifically address the issues related to VR, such as immersion, presence, spatial awareness, physical comfort, etc. In this chapter, we present relevant evaluation techniques for such experiences, and explore these identified measures through a user study.

The user study was designed to evaluate the effect of display type on the viewing experience. In doing so, we were also able to critique each measure’s ability to differentiate between display systems and to evaluate the CVR experience in general. This chapter describes that user study, and is reproduced from sections 3–6 of our paper published in the 2017 *Proceedings of the IEEE Virtual Reality Conference (IEEEVR)* [17].

## 5.1 Study design

A between-groups experiment of 63 participants was conducted, in which participants watched 360° videos in one of three display conditions. The details of this study are



**Figure 5.1:** A 360° video being watched on a head-mounted display (left), a TV (right), and our SurroundVideo+ system (centre)

described below. The study was approved by the UCL Research Ethics Committee (project ID 8923/001).

### 5.1.1 Subjects

Participants were recruited through a university mailing list and two participant pool websites. A total of 65 participants took part in the study, but data from two was excluded due to procedural issues. For one this was due to a hardware malfunction, while one participant decided not to watch one video containing bees for phobia-related reasons. No details were collected on the vision of participants due to an omission in the subject requirements, however no participants reported any difficulty with eyesight or completing any of the tasks due to vision related issues. Of the 63 remaining participants, 27 were male and 36 were female. Ages ranged from 19 to 76 years (mean: 27.78; standard deviation: 9.27). Participants were randomly assigned to a display condition until a condition reached 21 participants, at which point participants were randomly assigned between the remaining conditions. As a result, each display condition had 21 participants.

### 5.1.2 Experimental conditions

Each participant watched all videos in one of three display conditions, which are described below.

#### 5.1.2.1 Head-mounted display

The HMD used in this condition was the Oculus Rift CV1, driven by a Windows 8.1 desktop PC with an Intel i7-4790 CPU running at 3.6GHz with 8GB of RAM. The video card in use was a NVIDIA GeForce GTX 1080. The Oculus CV1 has a refresh rate of 90Hz. Whirligig was used as the video playback software. The videos played

with no visible lag at their expected frame rate of 30fps. Attempts were made to keep as many aspects as possible consistent across all three display conditions. To that end, in all display conditions the viewer was seated in the CAVE™-like display even when the displays were turned off, as was the case in the HMD condition. Audio across all three conditions was provided by the same stereo speakers, mounted above the corners of the walls of the CAVE™-like display. The built-in headphones of the Oculus Rift CV1 were removed, and the audio volume was the same across all conditions.

### 5.1.2.2 SurroundVideo+

The SV+ display is intended as a middle ground between the highly immersive visuals of a HMD – in that SV+ entirely fills the horizontal FOV of the viewer – and the directed experience of a TV. In Microsoft’s IllumiRoom display, peripherally projected content was designed to spatially augment a living room. In our SV+ display, we imagine a future scenario in which projection mapping can be used to visually negate the appearance of the surrounding room and replace it with the desired peripheral content.

The SV+ system used a three-walled CAVE™-like system to display the peripheral content, with a TV placed centrally to provide a focal point, as shown in the centre image in Figure 5.1. The TV was placed on a 70cm high table. Each wall of the CAVE™-like display was 3m long and 2.2m high, with each wall having a resolution of 1400x1050. While the CAVE™-like display in use also had a floor projector available, this was not utilised. The viewer sat on a fixed chair, positioned slightly back from the centre of the CAVE™-like display, meaning the peripheral projections covered the entire horizontal visual field of a viewer looking forward. No head tracking was used, as the videos do not support parallax, and head turning does not have an effect on the display.

The TV in use was a 60” LED HD TV made by LG, model number 60LA620V. The TV and projectors were driven by a Windows 7 desktop PC with an Intel i7-3930K CPU running at 3.2GHz with 32GB of RAM. The video card in use was a NVIDIA Quadro K5000. With a viewing distance of just over 2m, the TV subtended a solid angle of approximately 36° horizontally and 21° vertically for the head position of an average viewer. This meant the diagonal FOV of the TV was around 41°.

In order to allow 360° videos to be repurposed for use in the SV+ display, in a

pre-processing step, tracking data was created to ensure the content deemed important at any given time in the video was displayed centrally on the TV. This data was produced by manually annotating the videos with directional keyframes. Each keyframe specified a “forward” direction that indicated what the authors felt was the most important content in the viewing sphere at that moment. An example of this would be that, in general, a character was considered to be the most important scene element when they were speaking. These keyframes were then linearly interpolated to produce a “forward” direction per frame. This meant that, when played, the viewing sphere rotated, tilted and cut around the viewer to maintain the important content on the TV as it moved through the scene. This tracking data produced video playback with a clear narrative, however the authors are non-expert directors so the visual experience may not be optimal.

Video playback in the SV+ system was achieved using an adapted version of the open source OmiPlayer 360° video player written by Omar Mohamed Ali [116]. While the projectors and TV had high refresh rates of 96Hz and 200Hz respectively, the videos were only available with a frame rate of 30fps. The system was capable of displaying the videos at this frame rate with no visible lag.

### 5.1.2.3 Television

The TV display condition was identical to the SV+ display condition, except that the peripheral projections were disabled. For ease of swapping conditions, the projectors were not switched off but instead they projected solid black. For this reason, the walls of the CAVE™-like display during the TV condition were at the projectors’ black level.

## 5.1.3 Stimuli

Videos that matched the genre requirements were selected based on several factors. The videos needed to be suitably engaging, and high in audio and video quality. To ensure a contiguous image between the projected content and the TV, the FOV of the content displayed on the TV was fixed at the angle subtended by the physical TV within the CAVE™-like display. As discussed in section 5.1.2.2, the diagonal FOV of the content displayed on the TV was 41°. This meant the videos needed to match certain character placement requirements, and videos in which characters were too close to the camera had to be discarded.

Due to the small solid angle subtended by the TV, videos with as high a resolution as possible were required. Source videos were only available with a maximum resolution of 4K. Even with a 4096x2048 pixel equirectangular video, the effective resolution of the content displayed on the TV was only 409x250 pixels. It is fair to say that this is a noticeably low resolution for a TV capable of displaying up to 1080p. As the participant was seated just over 2m from the display, this did not seriously distract them.

Videos were also selected based on their appropriateness for the designed measures, which are discussed in full in section 5.1.5. For example, the spatial awareness task required participants to mark the locations of objects that had been seen in the video on a floor plan of the scene. This required videos that were staged largely in a single space, and had objects that could reasonably be remembered. The Measuring Narrative Engagement Questionnaire (MNEQ) required narrative content, which proved particularly difficult to source. This was due to the limited duration of content, as well as the fact that difficulties in storytelling mean many 360° videos are experiential in nature rather than narrative based.

It was decided that participants would sit in a fixed chair across all conditions. During an informal pilot study, participants in the HMD condition failed to observe any action happening behind them, including a dramatic fight sequence. As a result, content was reselected in which most action happened in the forward direction, and all critical action happened within  $\pm 100^\circ$  from “forward”. This may have had unintended consequences for some of our measures, as will be discussed later in section 5.2.

Four videos were selected. The first, a 2m19s music video called ‘Graffiti’ by Noa Neal, was used to reduce the novelty effect of the display. The DOCUMENTARY stimulus was a 5m53s documentary piece about a beekeeper by BBC R&D. The HORROR stimulus, which was of the slasher sub-genre, was ‘Hide and Seek’ by BriaAndChrissy and was 2m47s long. The final video, the NARRATIVE stimulus, was called ‘Real Memories’ and was made by MINI. This video told the story of a murder and lasted 4m43s. All of these videos are available online. The URLs for each are shown in Appendix B.

### 5.1.4 Hypotheses

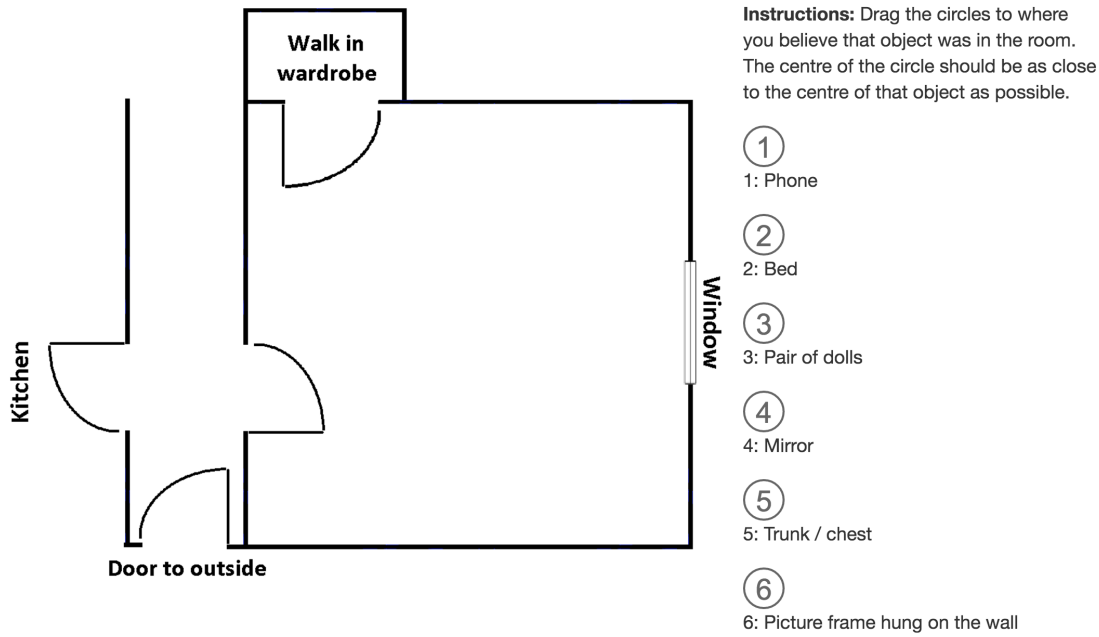
Hypotheses were based on previous work, as discussed in Section 2.3, with an emphasis on areas in which 360° media is likely to offer an improved experience or be at a disadvantage over traditional format media. Additionally, some hypotheses were based on feedback from industrial sources.

The following hypotheses were investigated:

- H1** There will be a difference in spatial awareness between conditions
- H2** There will be a difference in incidental memory between conditions
- H3** There will be a difference in narrative engagement between conditions
- H4** There will not be a difference in video enjoyment between conditions
- H5** There will be a difference in display enjoyment between conditions
- H6** Attention can be guided in TV/SV+ conditions
- H7** There will be a difference in participant's concern about missing something between conditions
- H8** Participants will be more afraid during a horror video in the HMD condition

### 5.1.5 Measures

There is a vast range of questions that can be asked about immersive media. We selected a subset of these measures so as not to overwhelm participants, and to keep each trial to under an hour to prevent fatigue. As CVR is a new field of study, our focus for the selection of metrics has been based on conversations with the CVR community – such as concerns that have been raised about the format by producers – as well as areas we believe are promising and applicable based on previous VR research. While there is little academic basis so far, open-ended, qualitative studies are emerging that indicate that these concerns are not just being felt by content producers, but are indeed critical aspects of the end-user experience [117]. Part of the selection process was determining what effective measures can be used, and what measures might be generally useful. A full list of background questionnaire questions, as well as questions for each stimulus, is included in Appendix C.



**Figure 5.2:** Object placement task for HORROR stimulus. Participants completed this task on a laptop computer, dragging the circles using the trackpad.

**H1: Spatial awareness** To measure spatial awareness, participants were asked to mark the locations of objects seen in the video on a floor plan of the room featured in the video. As two of the stimuli were set mostly within a single location per video (HORROR and NARRATIVE), the spatial awareness test was administered for these videos. The map placement task for the HORROR stimulus can be seen in Figure 5.2. Objects were represented as circles to ensure participants did not focus on orientation, and were labelled to ensure each placed object had a known corresponding object in the ground truth.

Objects were chosen at varying levels of difficulty, from items that characters interacted with that formed an element of the plot, to more difficult items that would be easy to miss. For reasons of fairness, all objects were visible in all display conditions. At least one participant in each condition correctly placed all objects.

The final measure was taken as the summed Euclidean distance of each placed object from a “ground truth” object placement. As the videos in question were not filmed by us, the “ground truth” and floor plans were to some extent approximations based on close inspection of the videos.

**H2: Incidental memory** To measure incidental memory, participants were asked ten questions about content from the DOCUMENTARY stimulus. All memory questions were taken from the audio track and had no visual reinforcements. While the video features a visible narrator, his face is masked by a beekeeper's veil and therefore no additional information is gained by looking at him. The audio was delivered through the same speakers and at the same volume across all three display conditions.

The questions varied in difficulty. The easiest was a fact that was said twice in different wordings, "There are two microphones on the honeycomb...I like to record stereo in the hive". Memory of this statement was checked with the question, "How many microphones were on the honeycomb?". The hardest questions related to difficult-to-remember concepts that were mentioned in passing, for example the statement "I'm not allergic to stings, which one in every thousand people are" for the question "What ratio of people are allergic to bee stings?".

**H3: Narrative engagement** Narrative engagement was measured using Busselle et al.'s Measuring Narrative Engagement Questionnaire (MNEQ) following the two narrative stimuli (HORROR, NARRATIVE). The MNEQ was not applied following the DOCUMENTARY stimulus, as this piece is non-narrative and many of the MNEQ questions would not make sense in this context.

**H4, H5: Enjoyment** Enjoyment was measured using five-point Likert scale indications of agreement with two statements. We wished to measure enjoyment of the display, rather than enjoyment of the video. In an attempt to tease apart these enjoyment levels, participants were asked to indicate their level of agreement with two statements: "Considering the display and the video separately, I enjoyed watching this video" and "Considering the display and the video separately, I enjoyed using this display". These two questions were placed side by side on the questionnaire to ensure participants answered them in tandem.

**H6: Attention guided** To test how successfully attention could be directed in the TV/SV+ conditions over the theoretically undirected HMD condition, elements of the video that could be examined via questionnaire were highlighted in the TV/SV+ conditions by centring them on the TV. For example, this included a lingering shot of the murder weapon prior to the murder in the NARRATIVE stimulus, and cuts to a telephone call being initiated by a victim in the HORROR stimulus. These were examined

using the questions “What was the murder weapon? Describe it as specifically as you can (colour, shape, material)”, and “Who initiated the phone call?” respectively. In total four such questions were asked. Three questions had pass/fail answers that contributed zero or one to the total attention score, and one was marked out of three depending on the level of detail provided of the murder weapon. This gives a total possible score range of zero to six. To ensure fairness, the answers were marked blind, i.e., the display condition of the participant was not known when their attention score was tallied.

**H7: Concern about missing something** Concern about missing something was measured by five-point Likert scale responses of agreement to the statements, “At times, I was worried I was missing something”, and, “My concern about missing something impacted my enjoyment of the video”. These questions were not placed side by side on the questionnaire. The responses to these two questions were summed together to give an indication of a participant’s general concern about missing something.

**H8: Fear during horror** To improve the validity of questionnaire responses, participant’s fear during the HORROR stimulus was measured using two questions. The question, “I felt afraid while watching this video”, was designed to measure fear directly. The question, “I felt nervous while watching this video”, was designed to measure anxiety, a state associated with fear during horror media [118]. Participants responded to both questions on a five-point Likert scale of agreement. These questions were not placed side by side on the questionnaire. These answers were summed to give a total score.

### 5.1.6 Procedure

Due to the learning effects that would exist in the metrics for memory and spatial awareness, a between-groups design was required. Each participant, therefore, watched all videos in one of the three display conditions. As many elements as possible outside of the display device were held constant, including the video order, audio and chair used.

Before arrival, participants were assigned to a display condition using a random number generator. Upon arrival, participants were given an information sheet to read. Important aspects of the information sheet were reinforced verbally, such as the procedure and the risk of simulator sickness. Participants then signed a consent form, and were introduced to the experiment environment.

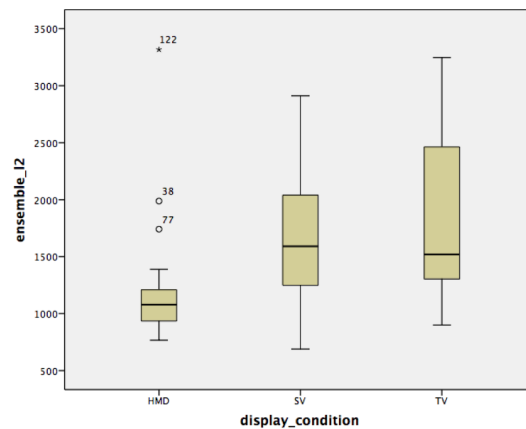
The experiment consisted of four videos. The first – a 2m19s music video – was designed to reduce the novelty effect of the display and was not followed by any questions. After each of the other three videos, participants were given a laptop to answer a set of questions immediately after the video’s conclusion. Before beginning the experiment, all participants were given identical advice about the nature of the questions, specifically that they would relate “to the content and their experience”.

As the DOCUMENTARY stimulus was the video used to measure incidental memory, it was viewed second. This was to reduce the likelihood that participants’ viewing of the video would be influenced by the questions for other metrics. For example, by knowing that questions regarding spatial awareness would be asked, it may have caused a participant to focus on remembering objects in the video, rather than viewing the video as naturally as possible in an experimental setup. At the end of the DOCUMENTARY stimulus, participants answered questions to test incidental memory, as well as generic questions about the experience such as their concern about missing something and enjoyment.

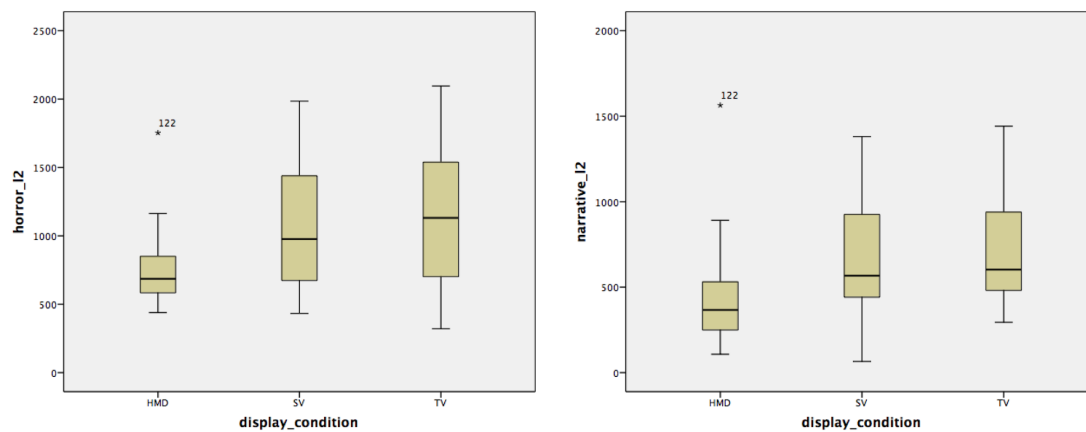
The HORROR stimulus was watched third. Following this video, participants answered questions related to their feelings of fear during the video, and questions that tested their memory of visual aspects that the TV and SV+ conditions deliberately drew attention to. Participants also completed the MNEQ. Generic questions such as enjoyment and concern about missing something were also answered. Finally, participants were asked to place objects from the video on a floor plan of the room to test spatial awareness.

The forth and final video was the NARRATIVE stimulus. As with the HORROR stimulus, participants answered questions on attention, narrative engagement, enjoyment and concern about missing something. Participants also completed a spatial awareness task. Participants then completed a SSQ. Participants then took part in a short, unstructured interview to gain insights into their reasoning and opinions.

After the experiment, participants were debriefed about the purpose of the study and given the opportunity to use the HMD if they were not assigned to the HMD condition during the experiment. Finally, participants were given £10 for taking part and dismissed.



**Figure 5.3:** Object placement task ensemble results.



HORROR stimulus results.

NARRATIVE stimulus results.

**Figure 5.4:** Object placement task results by stimulus.

## 5.2 Results and discussion

Due to the ordinal nature of the Likert, attention and memory data – and outliers in the Euclidean error distance data – analysis was conducted using the Kruskal-Wallis H Test for all hypotheses. Results were not similar for all groups, as assessed by visual inspection of boxplots of the data. The results of the Kruskal-Wallis H Tests can be seen in Table 5.1. As eight hypotheses were being testing, Bonferroni correction was applied where the statistical significance required was  $p < .00625$ .

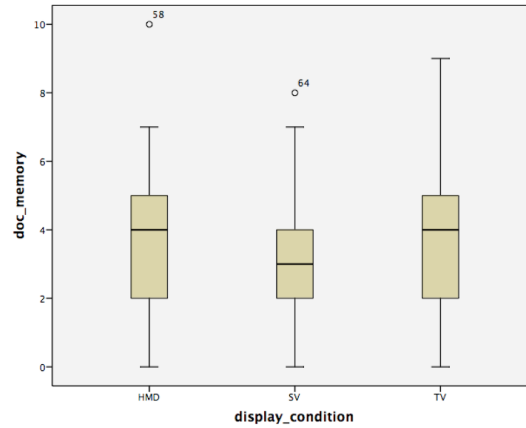
**H1: Spatial awareness** A boxplot of the ensemble spatial awareness results is shown in Figure 5.3, while breakdowns of the spatial awareness results by stimulus are shown in Figure 5.4. There was a statistically significant difference between display conditions in our measure of spatial awareness, as shown in Table 5.1. Pairwise comparisons were

**Table 5.1:** Results of Kruskal-Wallis H tests across all hypotheses.

Dependant variable	Stimulus	Mean rank			df	$\chi^2$	Asymp. Sig.
		HMD	SV+	TV			
Spatial awareness <sup>↓</sup>	Ensemble	19.90	36.19	39.90	2	14.146	.001* <sup>†</sup>
	HORROR	23.29	35.05	37.67	2	7.334	.026
	NARRATIVE	21.81	35.29	38.90	2	10.145	.006
Incidental memory	DOCUMENTARY	31.33	30.02	34.64	2	0.723	.697
Narrative engagement	Ensemble	35.17	32.50	28.33	2	1.485	0.476
	HORROR	34.93	31.62	29.45	2	.953	.621
	NARRATIVE	35.74	34.12	26.14	2	3.306	.191
Enjoyed video	Ensemble	37.69	29.36	28.95	2	3.152	0.207
	DOCUMENTARY	37.10	26.10	32.81	2	4.556	.102
	HORROR	32.81	35.50	27.69	2	2.363	.307
Enjoyed display	NARRATIVE	34.64	33.38	27.98	2	1.960	.375
	Ensemble	42.83	32.00	21.17	2	15.196	0.001* <sup>†</sup>
	DOCUMENTARY	40.67	31.81	23.52	2	10.383	.006
Attention	HORROR	40.69	35.21	20.10	2	16.378	.0003
	NARRATIVE	41.38	30.29	24.33	2	10.584	.005
	Ensemble	31.62	28.38	36.00	2	1.969	.374
Concern about missing something <sup>↓</sup>	HORROR	32.00	32.00	32.00	2	.000	1.000
	NARRATIVE	31.31	27.74	36.95	2	2.884	.236
	Ensemble	27.10	37.00	31.90	2	3.103	.212
Felt afraid	DOCUMENTARY	27.64	38.21	30.14	2	3.942	.139
	HORROR	32.40	33.02	30.57	2	.210	.900
	NARRATIVE	25.55	35.48	34.98	2	4.025	.134
	HORROR	28.14	31.55	36.31	2	2.155	.341

\* p &lt; .05

<sup>†</sup> p < .00625 (Bonferroni corrected for 8 hypotheses)<sup>↓</sup> A smaller value indicates a better result

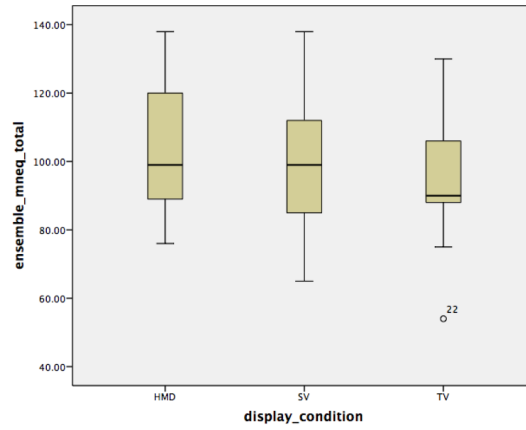


**Figure 5.5:** Boxplot of incidental memory results.

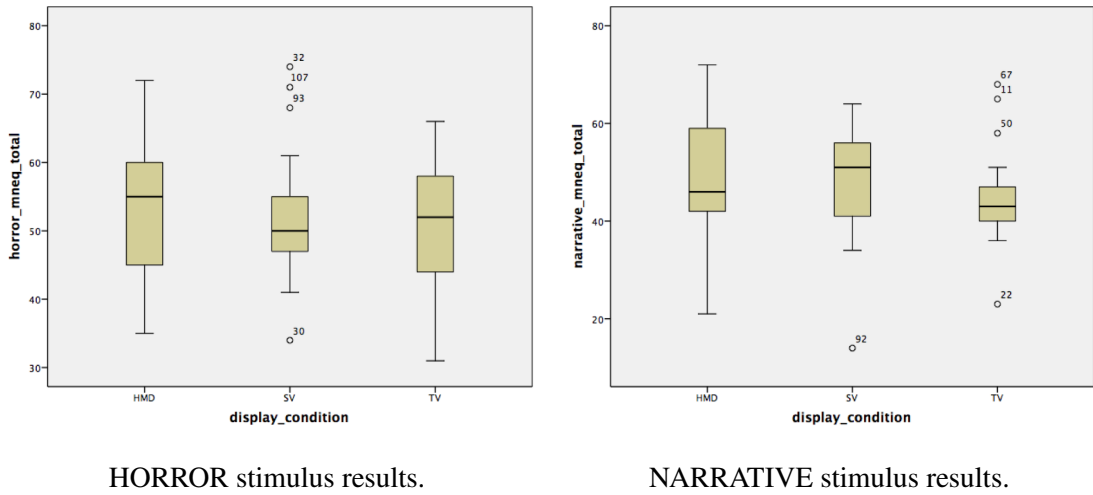
performed using Dunn’s procedure with a Bonferroni correction for multiple comparisons [119]. Adjusted p-values are presented. Values are mean ranks unless otherwise stated. This post hoc analysis revealed statistically significant differences in ensemble Euclidean error scores between the HMD (19.90) and SV+ (36.19) ( $p = .012$ ), and HMD and TV (39.90) ( $p = .001$ ) display conditions, but not between the TV and SV+ display conditions ( $p = 1.0$ ).

These results indicate that the HMD produced better spatial awareness than SV+ and TV displays. SV+ does not offer a statistically significant improvement to spatial awareness over the TV. This is somewhat unexpected, as the SV+ provides a horizontal FOV beyond that of human vision, while the TV offers only a  $32^\circ$  horizontal FOV. We propose two possible explanations for this. The first is that the key to producing good spatial awareness is exploration of the space, and the SV+ discourages exploration by providing a focal point. An alternative explanation is that rotation of the virtual space played a role. The HMD display was the only condition in which the virtual world is fixed with regards to the participant. In the SV+ and TV conditions, the world rotates and tilts to ensure the important content remains centred on the TV. It is possible that this rotation disoriented viewers, resulting in poorer spatial awareness. Further investigation is required to clarify these results.

**H2: Incidental memory** A boxplot of the memory scores is shown in Figure 5.5. No significant difference in incidental memory was found between conditions, as shown in Table 5.1. We were unable to recreate the results of [84], which indicated that incidental memory may be lower in a HMD over a tracked TV experience. Memory is highly



**Figure 5.6:** Boxplot of ensemble results from the MNEQ.

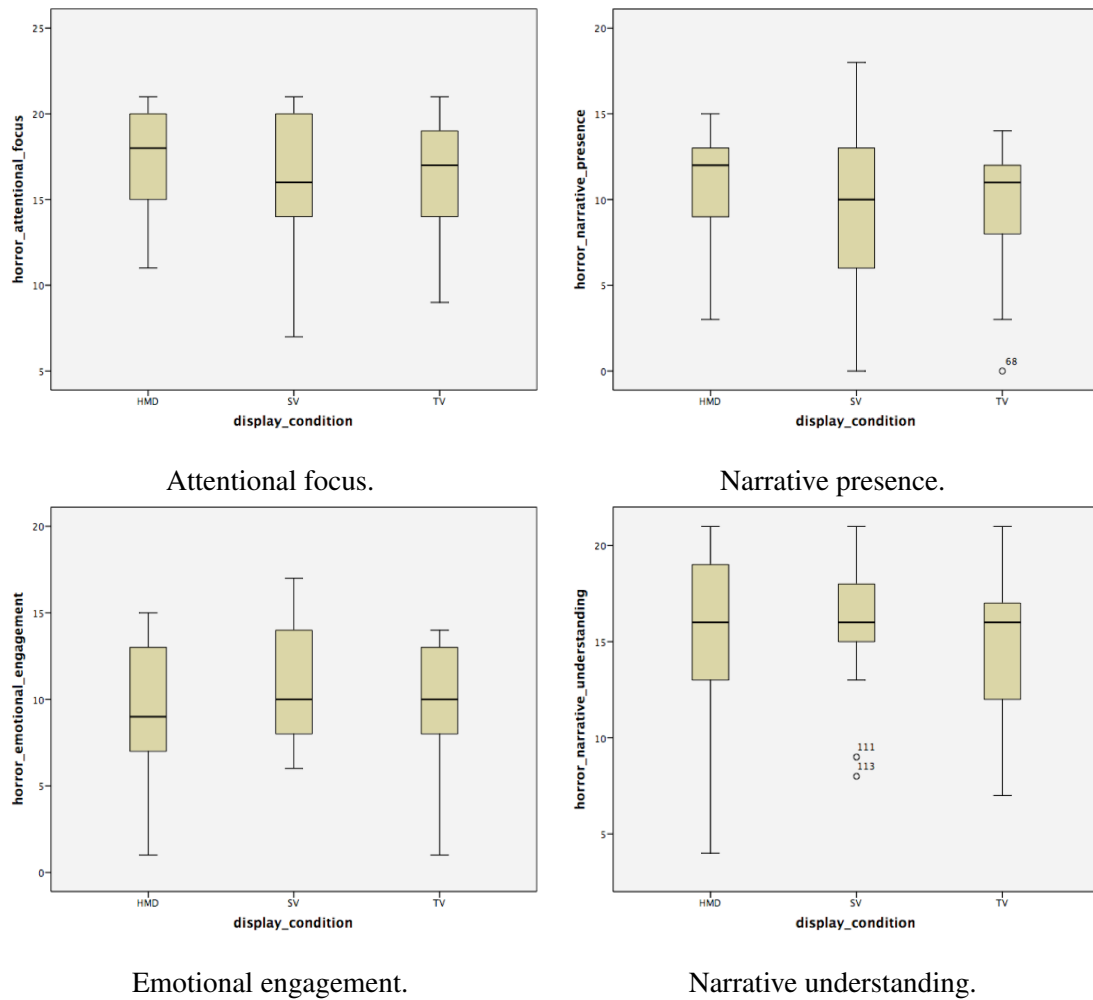


**Figure 5.7:** Results for the MNEQ by stimulus.

variable between individuals, however, and therefore more than 63 participants may be required to produce a statistically significant result, if such an effect exists.

**H3: Narrative engagement** A boxplot of the ensemble narrative engagement results is shown in Figure 5.6, and a breakdown by stimulus is shown in Figure 5.7. While the order of mean rank scores for ensemble narrative engagement decreased from HMD, to SV+ and then TV, the results were not statistically significant, as shown in Table 5.1.

The MNEQ can be broken down into four sub-scales: attentional focus; narrative presence, emotional engagement; narrative understanding. Boxplots for these sub-scales results for the HORROR stimulus are shown in Figure 5.8, while boxplots of the sub-scale results for the NARRATIVE stimulus are shown in Figure 5.9. Kruskal-Wallis H tests were used to test for statistically significant differences between display conditions for the MNEQ sub-scale results. None of the sub-scales revealed a statisti-



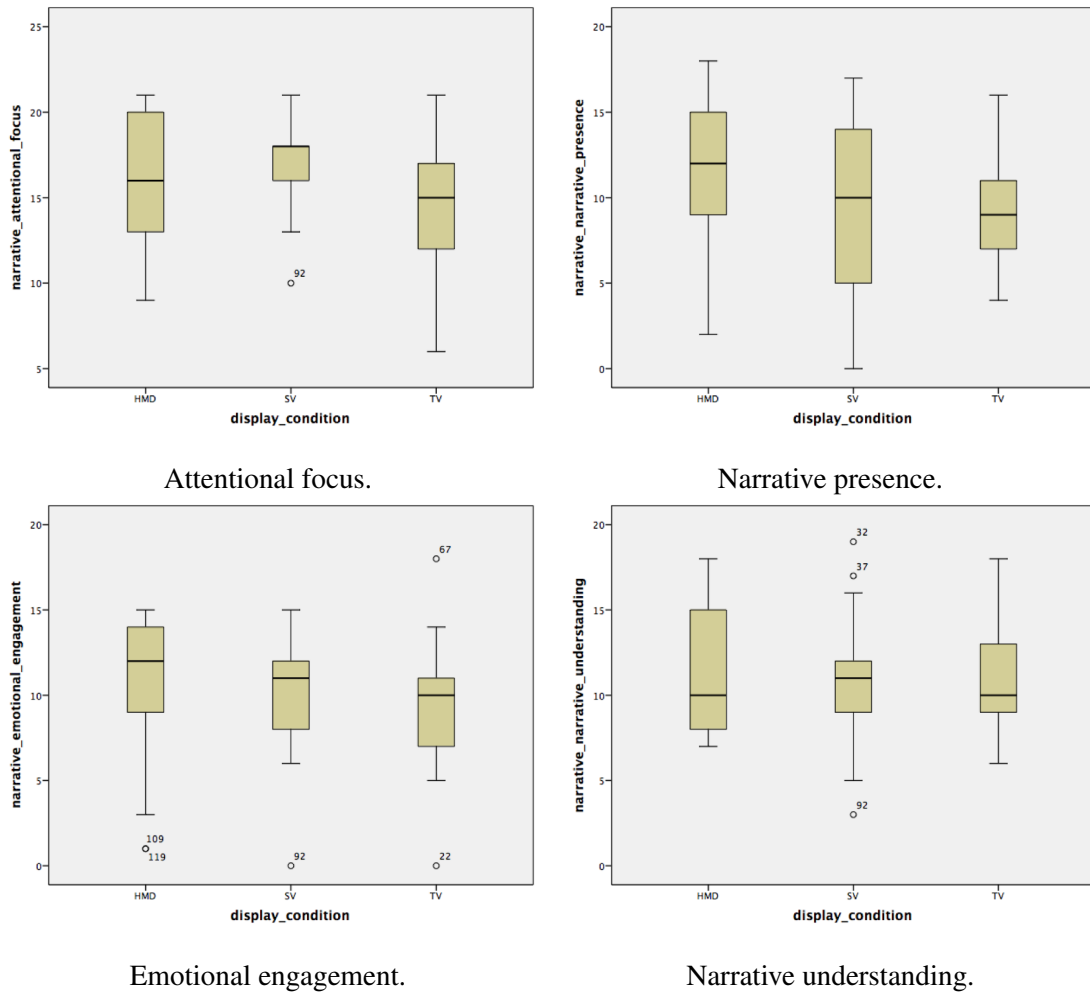
**Figure 5.8:** Boxplots for the four MNEQ sub-scales for the HORROR stimulus.

**Table 5.2:** Kruskal-Wallis H test results for MNEQ sub-scales for the HORROR stimulus.

MNEQ sub-scale	Mean rank			df	$\chi^2$	Asymp. Sig.
	HMD	SV+	TV			
Attentional focus	36.33	30.40	29.26	2	1.822	.402
Narrative presence	36.50	30.31	29.19	2	1.957	.376
Emotional engagement	30.00	35.52	30.48	2	1.183	.554
Narrative understanding	31.88	34.26	29.86	2	0.613	.736

cally significant difference for either stimulus, even before the application of Bonferroni correction. The results of these statistical analyses for the HORROR stimulus are shown in Table 5.2, while results for the NARRATIVE stimulus are shown in Table 5.3.

While this may indicate that narrative engagement is not strongly affected by display condition, it may also be a result of the short-form content that was used. While the short films used in the experiment were generally well received by participants, narrative engagement as measured by the MNEQ may require a more substantial and



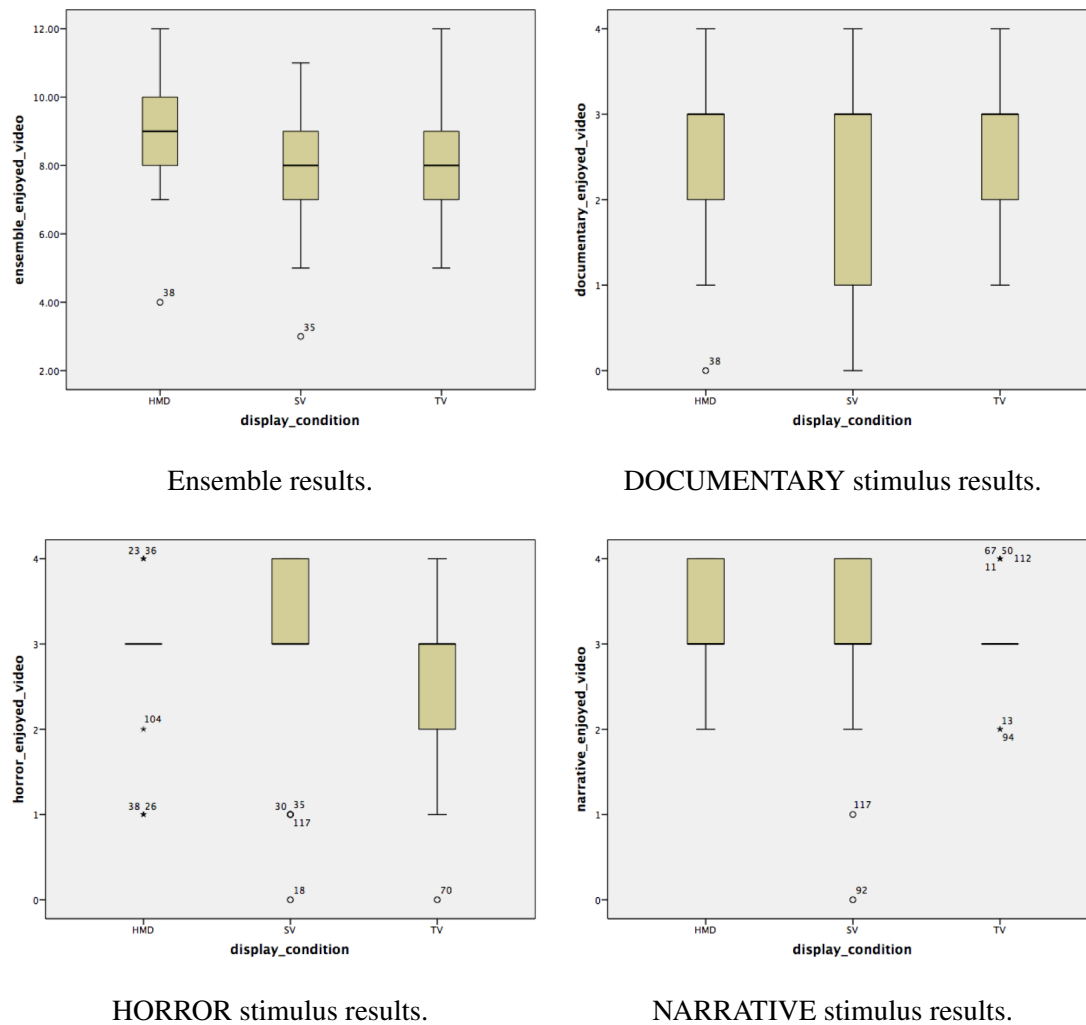
**Figure 5.9:** Boxplots for the four MNEQ sub-scales for the NARRATIVE stimulus.

**Table 5.3:** Kruskal-Wallis H test results for MNEQ sub-scales for the NARRATIVE stimulus.

MNEQ sub-scale	Mean rank			df	$\chi^2$	Asymp. Sig.
	HMD	SV+	TV			
Attentional focus	32.26	37.21	26.52	2	3.621	.164
Narrative presence	37.60	31.05	27.36	2	3.379	.185
Emotional engagement	36.21	32.67	27.12	2	2.654	.265
Narrative understanding	33.21	30.74	32.05	2	0.194	.908

engrossing plot than was offered in these clips. For example, questions such as, “During the program, when a main character succeeded, I felt happy, and when they suffered in some way, I felt sad”, may require a stronger connection with the characters than was achieved, as well as a more substantial narrative arc.

**H4, H5: Enjoyment** Boxplots of video enjoyment results are shown in Figure 5.10, while boxplots of display enjoyment are shown in Figure 5.11. A statistically significant difference between display conditions was present for display enjoyment, but not

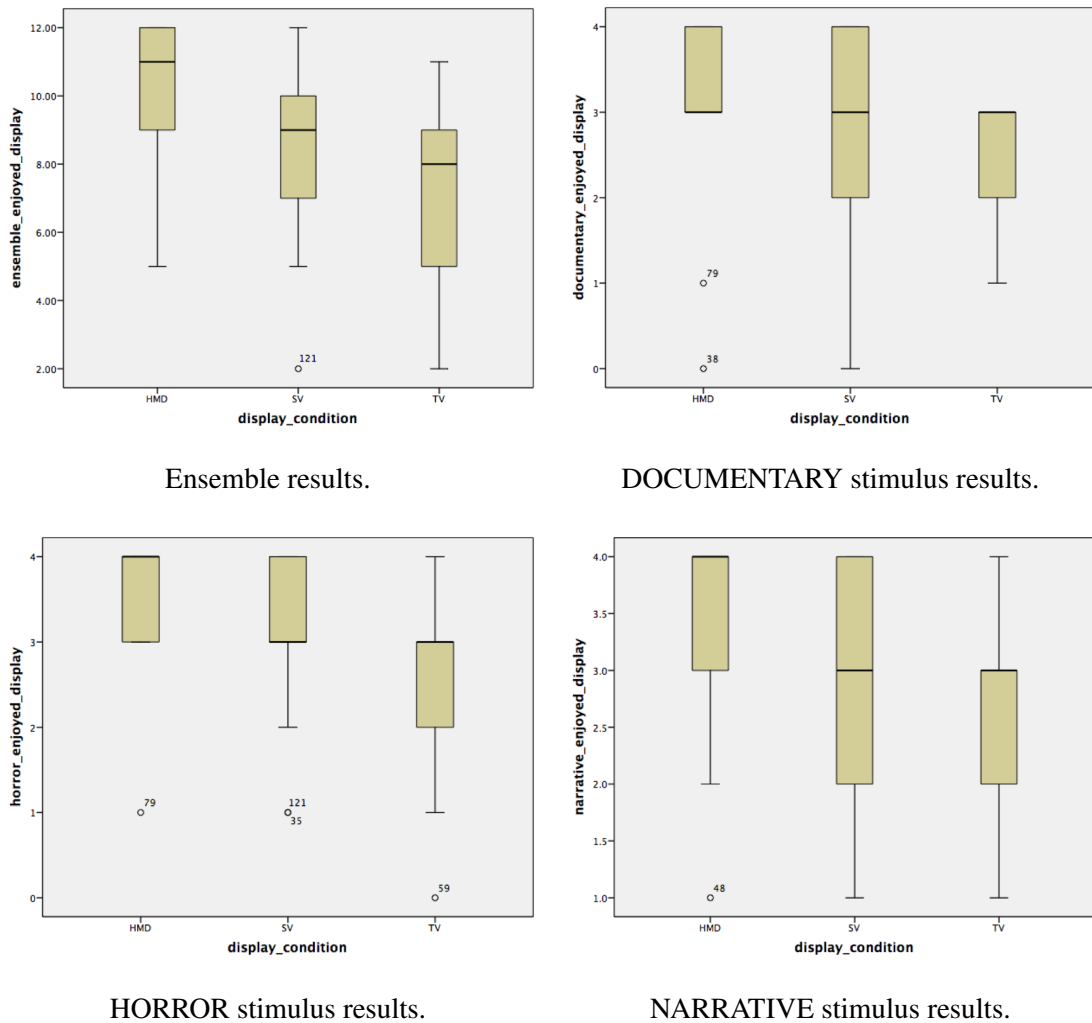


**Figure 5.10:** Boxplots for video enjoyment results.

for video enjoyment, as shown in Table 5.1. This indicates that participants were able to separate these concepts.

Pairwise comparisons of display enjoyment results were performed using Dunn's procedure with a Bonferroni correction for multiple comparisons [119]. Adjusted p-values are presented. Values are mean ranks unless otherwise stated. This post hoc analysis revealed statistically significant differences in ensemble display enjoyment scores between the HMD (42.83) and TV (21.167) ( $p = .000$ ), but not between HMD and SV+ (32.00) ( $p = .154$ ) display conditions, or between SV+ and TV display conditions ( $p = .154$ ).

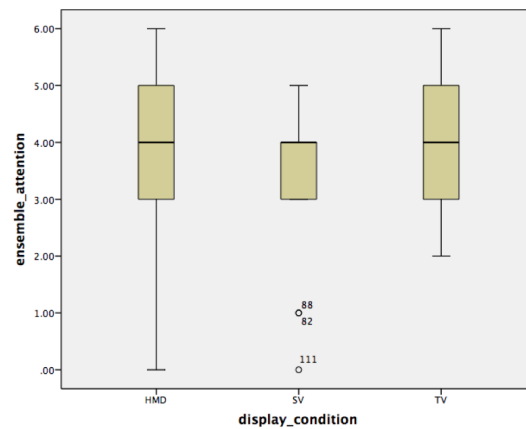
While the ensemble pairwise comparisons did not produce a statistically significant difference between the TV and SV+ conditions, pairwise comparisons do produce a statistically significant difference between TV and SV+ for the HORROR stimulus



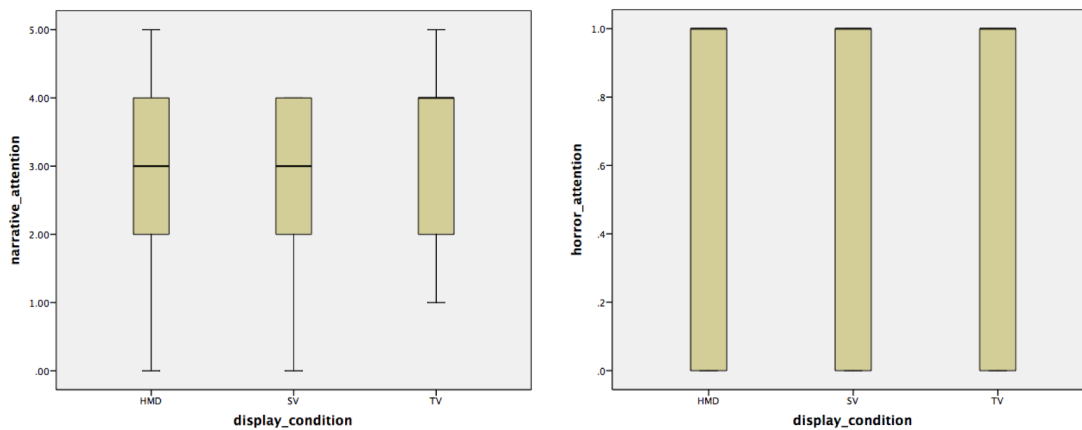
**Figure 5.11:** Boxplots for display enjoyment results.

( $p = .012$ ). While this must be considered a post hoc analysis, the difference in the level of display enjoyment between stimuli may indicate that certain types of content are more appropriate for a given display. Further investigation would be required to determine what may have made the HORROR stimulus particularly well suited for the SV+ display. Possible aspects that may have contributed include the genre, character placement, the rapidity of cuts required, and the characteristics of the captured environment.

The resolution of the TV may have had an impact on this metric. During the unstructured interview at the end of the experiment, several participants mentioned that the low effective resolution of the TV had impacted their enjoyment of the display. While the central content had an identical resolution in the TV and SV+ conditions, it is possible that the highly immersive visuals of the CAVE<sup>TM</sup>-like display partially



**Figure 5.12:** Boxplot of ensemble attention results.



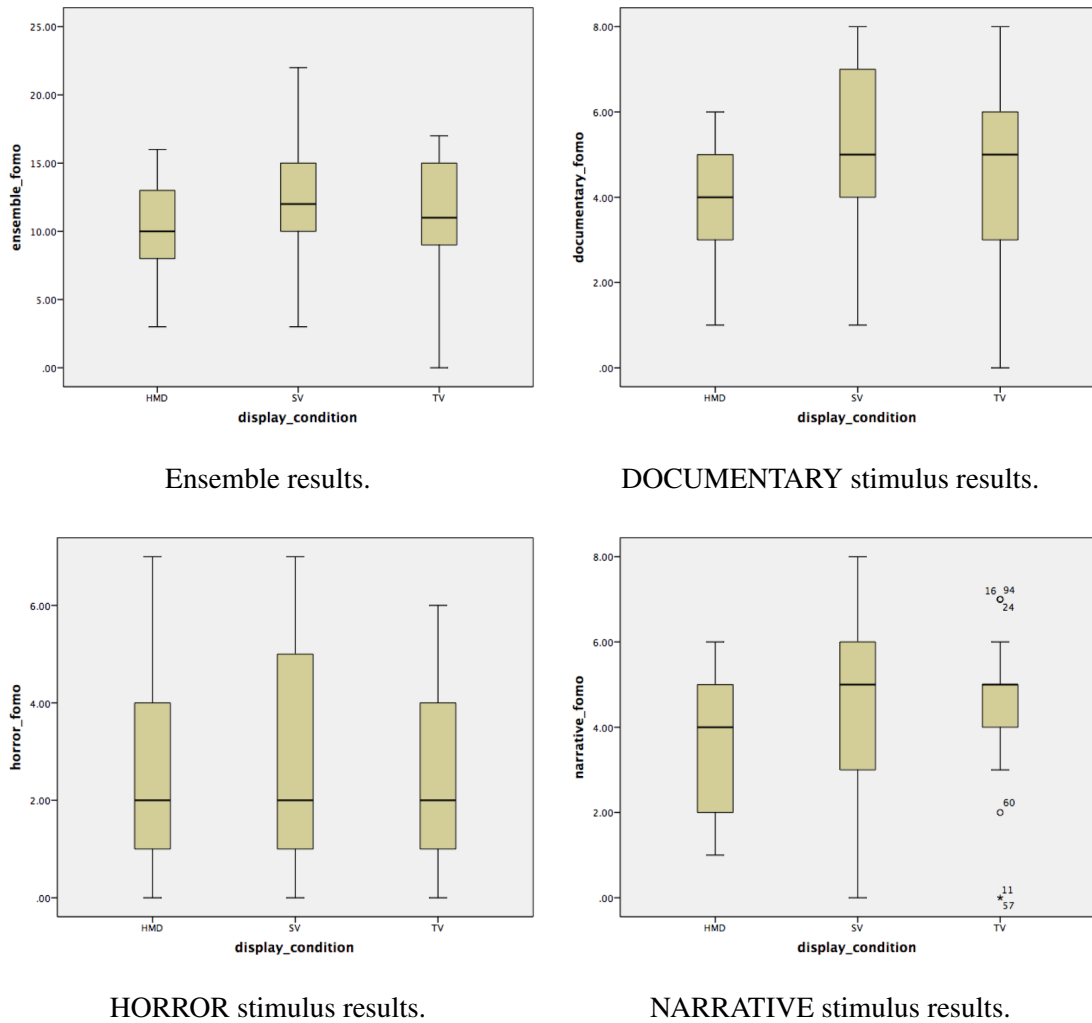
NARRATIVE stimulus results.

HORROR stimulus results.

**Figure 5.13:** Attention results by stimulus.

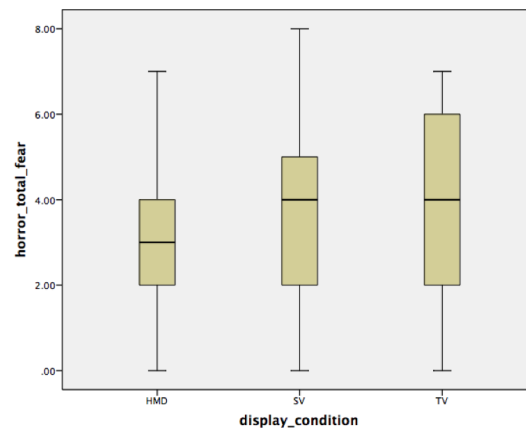
compensated for this issue in the SV+ condition, resulting in a stronger enjoyment result.

**H6: Attention guided** A boxplot of the ensemble attention results is shown in Figure 5.12, and a breakdown by stimulus is shown in Figure 5.13. No significant difference in attention was found between conditions, as shown in Table 5.1. This was an unexpected result, as it seems almost certain that a TV can draw attention to a specific aspect of the scene more effectively than a HMD. We propose two possible explanations for this. The first is that the chosen metrics for attention were not suitable, in which case a repeat of this study using different videos and attention measures may produce a statistically significant result. Another explanation is that the fixed chair used across all conditions limited the field of regard (FOR) of the HMD such that the risk of a viewer missing some aspect was substantially reduced.



**Figure 5.14:** Boxplots for users' concern about missing something.

As mentioned earlier, during a pilot study a video featuring a fight scene was replaced as important material was missed in the HMD condition. This decision was made to encourage fairness between the conditions. In the removed video, a character starts in front of the viewer, at roughly  $0^\circ$  from “forward”. This character then walks through the scene and becomes engaged in a fight directly behind the viewer, at roughly  $180^\circ$  from “forward”. No viewers were willing to crane in their seat to follow this character, despite his prominence, and became confused by the audio of a fight they could not locate in the scene. This may indicate that the FOR of a HMD when watched in the fixed chair is more limited than the available FOV. This may have interesting practical implications for current  $360^\circ$  content creation, as home viewers will most likely be seated on a fixed chair or couch.

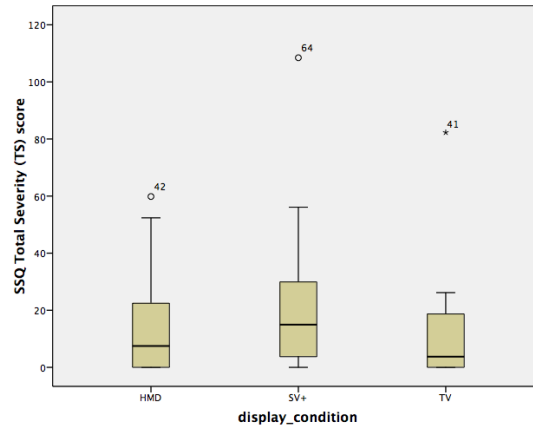


**Figure 5.15:** Boxplot of results for fear during horror.

**H7: Concern about missing something** Boxplots for a user’s concern about missing something are shown in Figure 5.14. In none of the videos did the HMD receive the highest mean rank score for concern about missing something, although the difference between display conditions was not significant as shown in Table 5.1. This result was unexpected, as there is a prevailing belief that HMD experiences often leave viewers feeling they have missed something. As discussed in the Attention results section, this may be because of the partially restricted FOR caused by the non-swivel chair, as well as the intentional choice of videos in which all action happens within  $\pm 100^\circ$  of forward. Due to this soft limit on the FOR, the HMD condition might also be considered partially guided.

The SV+ display received the highest mean rank score across all stimuli for this measure. This result is also unexpected, as the SV+ was intended as a partially guided experience. This may have been due to the feeling that viewers should be looking at the TV, but were aware that additional content was available in the periphery that they were unable to digest. These findings may also be a result of viewers’ subtle awareness that the content was designed for a  $360^\circ$  viewing experience, and that only a portion of the FOV was available in the TV and SV+ conditions.

**H8: Fear during horror** A boxplot of the scores for fear during the HORROR stimulus is shown in Figure 5.15. Displays with higher immersion produced lower levels of fear during the horror video, although not significantly as shown in Table 5.1. These unexpected scores may have been a result of the deliberate decision – due to concern about upsetting participants – to use a horror video that was extremely mild. One



**Figure 5.16:** Boxplot of SSQ total severity score.

participant reported finding the video “lame”, while none answered “Strongly agree” to the question “During this video, I felt afraid”. The type of horror – the slasher sub-genre – perhaps does not make best use of the characteristics of the HMD, such as being isolated, feeling physically vulnerable, and being entirely surrounded by the content. More psychological horrors, or horrors that rely on jump scares, might prove more effective in producing stronger differences between the display conditions. Worthy of note is that a participant in the HMD condition – who appeared visibly shaken by the video – reported in the post-experiment interview that the audio had played a key role in eliciting their fear.

**Simulator sickness** SSQ Total Severity (TS) scores were calculated using the formula specified by Kennedy et al. [108]. A boxplot of TS scores is shown in Figure 5.16. A Kruskal-Wallis H test was conducted to determine if there were differences in SSQ scores between the three display conditions: HMD ( $n = 21$ ), SV+ ( $n = 21$ ), and TV ( $n = 21$ ). Values are mean ranks unless otherwise stated. Distributions of SSQ scores were not similar for all groups, as assessed by visual inspection of the boxplot. SSQ scores increased from TV (28.14), to HMD (31.50), to SV+ (36.36) display conditions, but the differences were not statistically significant,  $\chi^2(2) = 2.192$ ,  $p = .334$ .

## 5.3 Limitations

Some issues were encountered when repurposing 360° videos that were designed for HMD viewing for display in the TV and SV+ conditions. Finding content that worked in all three display conditions, and also met our genre and quality requirements, meant

some compromises were required. This meant that some videos were not ideal for viewing in the SV+ condition. For example, the NARRATIVE stimulus featured conversations between two characters who were not standing near each other. To keep the speaking character on the TV required several cuts in quick succession. This was not a problem in the TV condition, as this is a standard visual technique used in TV. Some participants in the SV+ condition – who could see the other character in their peripheral vision – found this cutting irritating, prompting one to comment the display seemed to “flit about”. If the content had been filmed specifically for the purpose of this test or these displays, care could have been taken in character placement to ensure both characters could be framed centrally on the TV during conversations.

Our measure of spatial awareness may have been susceptible to a confounding factor, in that rotation of the virtual world with respect to the viewer was only present in two of the three display conditions (SV+ and TV). To clarify that the significant benefit in spatial awareness offered by the HMD was not caused by this rotation, an experiment should be conducted that investigates the relationship between exploration of the space and spatial awareness, with rotation controlled as an independent variable.

Rotations of the scene were generally well received. Some participants in the SV+ condition, however, did comment that the rotations had felt somewhat intense. This is to be expected, as the CAVE™-like display entirely filled the horizontal FOV of participants. No participants reported feeling unwell, and there was no significant difference in SSQ scores between display conditions. Our videos were all of short duration, however, so the risk of cumulative effects caused by the increased FOV for longer experiences has not been addressed.

A number of participants also commented on the low resolution in the TV condition. This is also to be expected, as the effective resolution of the TV was only 409x250 due to the solid angle subtended by the TV inside the 4k viewing sphere. Interestingly, participants in the SV+ condition tended not to comment on the resolution of the TV, despite it being identical to the TV condition. This may indicate that the immersive visuals provided by the CAVE™-like display compensated for the poor effective resolution of the TV to some degree. This low resolution was again caused by the decision to repurpose 360° video designed for HMD experiences. If content was to be shot specifically for the SV+ display, it is likely a similar approach to the original Surround

Video system would be used: capturing a high-resolution inset using a dedicated camera, while using separate cameras to capture the lower resolution peripheral content. This again emphasises that immersive video may not be a single type of experience: producers need to be aware of the type of display viewers will use. There may not be a single form of immersive video and guidelines for producers will need to recognise this.

Several participants commented on the audio mix in the DOCUMENTARY stimulus. Background noise and music present in the audio track may have hindered a participant's ability to hear the dialogue. This may have impacted their ability to remember the content, and therefore affected the memory metric. This was not a confounding factor, however, as all participants received identical audio across conditions.

Other improvements to the testing method include using longer format content when measuring narrative engagement. Longer format content may create a stronger feeling of engagement that may be more readily measured by the MNEQ, and thus may provide further insight as to if the display has an impact on narrative engagement. Additionally, using a more appropriate horror stimulus to test fear, for example one that uses jump scares, might elicit a stronger fear response – although such an investigation would need to be done with caution. Using content that was specifically filmed to match the requirements of our displays, for example aspects such as resolution and character placements, may provide a more accurate indication of the full potential of each display, as the compromises made to ensure repurposed content worked across all displays may have negatively impacted the results for certain metrics such as display enjoyment.

## 5.4 Conclusion

In this chapter, several metrics for measuring passive CVR experiences were discussed. These metrics related to areas in which CVR was likely to present an improvement over traditional film and TV experiences, as well as areas of concern for CVR that may impact its adoption. While there are many factors that are important to CVR experiences, such as display comfort, sociability, etc., it was important to limit the scope of our study to prevent participant fatigue. After careful review of the most appropriate and generalisable metrics, aspects considered measured spatial awareness,

narrative engagement, enjoyment, memory, fear, guiding attention, and the feeling of missing something.

A between-groups experiment was conducted with 63 participants. Three display conditions were investigated, including the radically different immersive displays of the HMD and the SV+, and a non-immersive TV condition. Our results indicated the HMD offered a significant benefit in terms of spatial awareness over both the TV and SV+ conditions. While it was expected that the TV would perform poorly on the spatial awareness metric, the SV+ was expected to perform better due to the large FOV provided. While this may have been caused by a lack of exploration encourage by the focal point of the TV, more experimentation is required to rule out the possibility of confounding factors. Significant improvements for enjoyment were present in the HMD over the TV and SV+, as well as the SV+ over the TV for the HORROR stimulus.

We were unable to confirm the work of a previous study that showed incidental memory may be lower in a HMD over a TV. We did not find a significant difference in narrative engagement between conditions, however this may be a result of the short-form stimuli that were available. The lack of a significant difference between displays in our measurement of fear during the HORROR stimulus may have been caused by the choice of an inappropriate video. The stimulus was not scary, and did not make use of the characteristics of the immersive displays likely to increase fear.

Drawing attention and a viewer's concern about missing something were also not significantly different between display conditions. These are unexpected results, as there is a commonly held belief that the undirected experience of the HMD causes viewers to miss visual events, as well as to experience a feeling of missing something. Our results may indicate that using a fixed chair in a HMD experience places a soft limit on the field of regard. This result would have important consequences for video production, as at-home viewers will likely be seated on a couch. Further experimentation is required to confirm this result, and to identify the FOR available for use in videos designed for consumption in the home.

It is clear that passive media viewing consists of a complex interplay of factors that present many challenges for evaluation. As CVR is becoming a common use for virtual reality hardware, it is essential for the virtual reality community to investigate this emerging field. Overall, one important steer for future research is the difficulty of find-

ing content for each display. The affordances of the displays are actually quite different, and thus content developed for one may not work on another. We have highlighted that comparing SV+ to HMD is difficult. Even evaluating different types of HMD content may be challenging, because of assumptions made on whether the participant can turn freely or not. In future work, we hope to explore further this complex area and cover a wider range of scenarios, such as videos with camera motion and social viewing of videos.

One particular area of interest for 360° media is the transportive nature of such experiences. Despite the fact that there is a prevailing belief that 360° media creates a form of presence, where the viewer feels transported to the space depicted in the scene and perhaps even involved in the action, measuring such a response presents many challenges. While measuring presence remains a challenge for the VR community at large, it is perhaps particularly difficult in 360° media experiences, where a lack of agency may disrupt proxies often considered to indicate presence, such as response-as-if-real [74]. Certainly a more concrete understanding of this phenomenon, as well as ways of measuring it, would be beneficial for the creation and evaluation of 360° media experiences.

The work presented in this chapter represents one of the first academic investigations of how to evaluate CVR content. By looking at the impact of various different kinds of display when viewing 360° videos, we have also broached the broader question of what techniques and tools are most applicable to the evaluation of 360° media experiences in general. Additionally, we have helped illustrate some of the likely pitfalls investigators must consider when designing experiments exploring 360° media experiences.

## Chapter 6

# User study: The effect of transition type in multi-view 360° media

Currently, captured 360° media is generally fixed viewpoint i.e. only the three degrees of freedom associated with orientation are available to the user to explore the scene. While free-viewpoint 360° media is being investigated (e.g. [26, 120]), there are still many technical challenges to be overcome. Multi-view 360° media (MV360M) – in which 360° views are captured from multiple locations – may offer a partial solution to the issues associated with fixed-viewpoint 360° media. This is achieved by allowing users to view the space from multiple perspectives. Systems have existed for some time that allow the exploration of spaces by transitioning between 360° images [14, 15], and video-based MV360M experiences are becoming more common [28, 29].

In MV360M systems, a visual effect is required to transition the user from one camera location to another. This effect may have a significant impact on user experience. While multiple transitions are available from standard film production, such as wipe, dissolve, fade, etc., there are aspects inherent in immersive MV360M that require special consideration. A fade to black in an immersive display, for example, is the equivalent of the world suddenly going dark, and with no visual features available, this transition might result in discomfort or disorientation.

Likewise, while work has been done to explore the impact of transition types in immersive experiences – particularly in the VR locomotion literature – certain attributes of MV360M make such systems inherently different from most real-time rendered experiences. For example, as each view in MV360M requires a physical camera to have been placed at that location, such locations tend to be limited in number. The limited

number and fixed-position nature of the views available may have a detrimental effect on a user's ability to understand the scene. This may mean that additional benefit could be gained from transition types that provide further information about the spatial layout of the environment over the same transition in real-time rendered experiences.

In this chapter, the impact of the transition type when viewing image-based MV360M is explored. This chapter is adapted from a paper published in *IEEE Transactions on Visualization and Computer Graphics* (TVCG). A repeated-measures experiment was conducted with 31 participants. Wearing a HMD, participants navigated through four static scenes, initiating transitions using a tracked hand controller. Three transition types were examined: *teleport*, *model* and *Möbius*. The teleport transition moves the user instantaneously to their selected location while leaving their orientation unaltered. The model transition moves the user linearly through a reconstructed 3D model of the scene. While parallax cues inherent in the model transition are likely to provide users with the most complete impression of the scene, such models are expensive and labour intensive to produce. The Möbius transition is proposed as a possible middle ground between the teleport and model transitions. The Möbius transition is an image-based transformation that gives the impression of movement between panoramas using a zoom effect. The motion cues provided by this effect may help to improve understanding of the scene and the transition, while as an image-based transition there is no requirement for a 3D reconstruction of the scene to be available.

The metrics investigated were spatial awareness, users' movement profiles, transition preference and the subjective feelings of moving through the space, disorientation, dizziness, and naturalness. Our results indicate that trade-offs between transitions will require content creators to think carefully about what aspects they consider to be most important when producing MV360M experiences.

## 6.1 Experimental Design

We conducted a repeated-measures user study to evaluate the effect of transition type when exploring a scene captured in 360° from multiple locations. At each location, participants could look around naturally via a tracked HMD, using the three degrees of freedom associated with orientation. Participants could move around inside the scenes by selecting different camera locations using a position-tracked, hand-held input de-

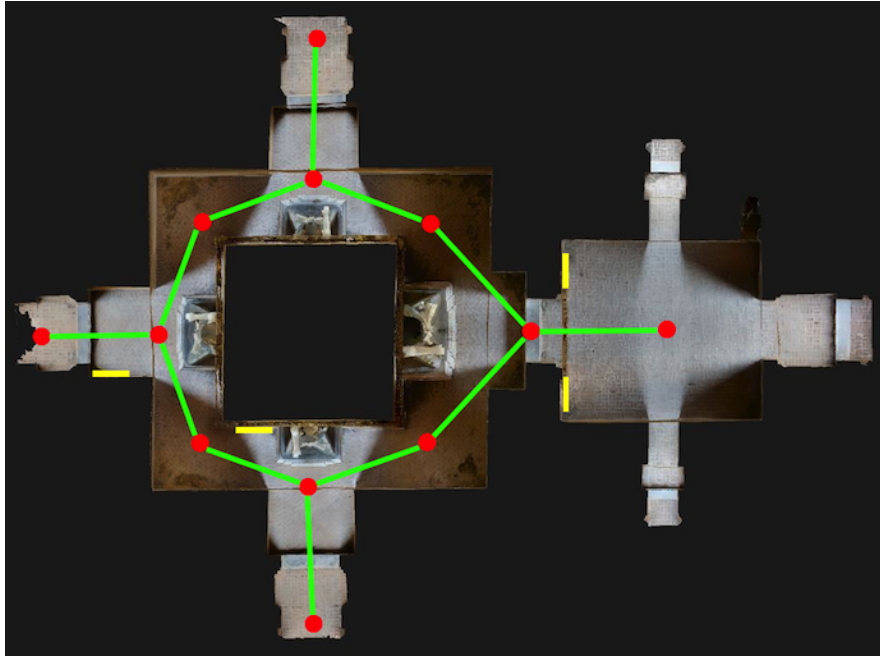
vice. While there were multiple buttons and triggers on the input device, they all performed the same action and users were free to use whichever button felt most comfortable. We did not collect data on which buttons participants used. When a user chose to move to another location, they were transitioned from their current location to their selected location via one of three transition types, each of which is described in section 6.1.2. Generally, the transition type was selected randomly by the system. The only time when the transition type was not randomly selected was before the pointing task or preference questions; at these times the transition type was counterbalanced. These tasks and questions are described fully in section 6.1.3, while the counterbalancing strategy is described in section 6.1.6.

### 6.1.1 Stimuli

As will be discussed in section 6.1.2, the model transition required a 3D reconstruction of the scene. As a result of this, the available stimuli was limited to static scenes. The stimuli was provided by Matterport, a VR capture company whose cameras record RGB-D data in 360°. Using this data, Matterport construct a textured 3D model of the scene. This provides 4k 360° images from set locations, as well as a 3D model of the scene. Matterport kindly provided this data to us for free for academic use. The quality of these 3D models is consistent with structured light scanning, i.e. they contain some holes and do not accurately capture fine details. The 3D model for one of our stimuli is available to view online [121].

In total, three models were used: a gallery, a Buddhist temple, and a house. As the house was set over two floors, this model was split into two independent scenes. This provides a total of four stimuli: *gallery*, *temple*, *house upstairs* and *house downstairs*. A map of the temple scene is shown in Figure 6.1, while statistics for all stimuli are shown in Table 6.1.

The stimuli was limited to indoor scenes. In general, scenes were highly occluded, meaning only a small number of cameras were in the line-of-sight of any other camera. As our setup only allows transitions to cameras that are in the line-of-sight of the current camera location, this required participants to navigate around the scene, transitioning from camera to camera, in order to complete the required tasks. This procedure is described fully in section 6.1.6. Camera locations were limited to the minimum number



**Figure 6.1:** A map of the temple stimulus. Camera locations are shown as red nodes, while green edges indicate available transitions between locations. Yellow lines show the locations of targets in this scene.

**Table 6.1:** Stimuli statistics. Stimuli are listed in the order they were presented to participants.

Stimulus	Camera locations	Transition edges	Targets
Gallery	4	3	2
House upstairs	11	10	6
Temple	12	12	4
House downstairs	11	12	6

possible, such that the scene was covered by a connected graph of transitions.

### 6.1.2 Transition Types

Three transition types were used: teleport, model and Möbius. The teleport transition was instantaneous, while both the model and Möbius transitions took six seconds to complete. A transition duration of six seconds was chosen to ensure transitions felt comfortable to participants. Each of these transitions is now described in detail.

**Teleport:** When a participant chose to move to a new location, and the transition was determined by the system to be a teleport, the participant was moved instantaneously from their original location to their selected location. On arrival at this new location, their orientation in the virtual space was the same as it had been at their start location. To achieve this, the image the user was seeing was instantaneously swapped from the panorama captured at their original location to the panorama captured at their selected location. Then, as in the other transitions, scene elements such as the available

locations to move to next were updated to be consistent with the new panorama.

**Model:** In the model transition, the user moves through a 3D model of the scene. This transition, therefore, requires a reconstructed 3D model of the scene. While movement easing types were explored, simulator sickness appears least severe when there is minimal changes in velocity [122]. Following advice from Oculus that it is the duration of velocity change that should be minimised, a linear movement with infinite acceleration and deceleration was used [123]. As the user sees a panoramic image when not moving between locations, a linear interpolation was used to blend between the panoramic images and the 3D reconstruction of the scene. The blend was necessary as the 3D model was not completely consistent with the panoramic images, for example the lighting was often noticeably different. This would likely always be the case for MV360M. Even if all cameras could have matching settings such as exposure, the location of the camera impacts aspects such as specular highlights, so the lighting would not be consistent between 360° cameras capturing the same scene. Blending between the panorama of the current location and the model lasted 0.5s, the linear movement from the current location through the 3D model to the new location lasted 5s, and the blend from the model to the panorama of the new location lasted 0.5s, resulting in a total transition time of 6s. Frames from this transition type can be seen in Figure 6.2.

**Möbius:** The Möbius transition is an image-based transition, in which a transformation is applied to give the impression that the user is moving from the panorama of the current location to the panorama of the next locations. This is achieved by “zooming in” to both panoramic images. While zooming in is a common technique in standard format media production that gives the illusion of getting closer to something, its application in panoramic media is complicated by the spherical nature of the imagery. In order to zoom in to a specific point in a spherical image, the rest of the content cannot be cropped, but must instead be compacted towards the zoom’s antipodal point.

The Möbius transformation was proposed by eleVR and Henry Segerman as a technique to allow zooming for panoramic media [124]. The Möbius transformation is a conformal mapping, in that it preserves local angles (for a review, see [125]). It can be used to enlarge the image towards the zoom point, while reducing the size of elements towards the zoom’s antipodal point. This technique can be used to give the impression of moving towards the zoom point, although deformations to the space mean the effect



**Figure 6.2:** The 3D model transition. First, the panoramic image for the original location (frame a) is faded out, revealing the 3D model (frame b). The user is then moved linearly through the scene (frame c). Finally, the panoramic image for the new location is faded in (frame d).

does not appear natural.

There are many possible ways to incorporate the Möbius transformation into a transition. In our implementation, the user selects a new location to move to. We call the panoramic image for the current location  $L_c$ , and the panoramic image for the selected next location as  $L_n$ . If a ray was cast from the current location to the new location in 3D space, the point of intersection on the viewing sphere of  $L_c$  becomes the “zooming point” for the Möbius transformation, referred to as  $P_{zc}$  here. Likewise, the zooming point for  $L_n$  is where the same ray would intersect  $L_n$ , referred to as  $P_{zn}$ .

First, a “zoomed out” version of  $L_n$  expands as a circle at  $P_{zc}$  until it reaches approximately 38% of the height of  $L_c$  when viewed in equirectangular form. A value of 38% was chosen for aesthetic reasons. The Möbius transformation is then applied to “zoom into”  $L_c$  and the zoomed out version of  $L_n$ . Assuming that the viewer is facing towards the new location, this means  $L_c$  collapses behind the viewer, while  $L_n$  expands over them. In our implementation, all transitions were generated as video files in a preprocessing step. This video file was then played back when a transition was

initiated, using the PopMovie video plugin for Unity. Due to larger videos causing unacceptable levels of lag in the rendering, the videos were downscaled and played at a resolution of 2048x1024. During a transition, the video was started and blended in over 0.5s, continued playing for 5s, and then blended out over 0.5s, resulting in a total transition time of 6s. This effect can be seen in Figure 6.3.

The effect produced by the Möbius transformation is difficult to describe accurately, and we would encourage readers to watch our videos of the transition in practice. We have recorded a complete user journey through the gallery stimulus [126], and made available a 360° video showing two Möbius transitions in the temple stimulus [127]. The code used to generate our Möbius transitions is available online [128].

### 6.1.3 Hypotheses

#### 6.1.3.1 Spatial Awareness

As in the related work discussed in section 2.3.2.1, spatial awareness was measured using a pointing task. Participants were asked to point at a known location in the scene that was no longer visible using a tracked hand controller. We refer to this task as the “pointing task”. Both the error angle and the time to complete the task were examined. The error angle was defined as the angle between the user’s pointing ray and the ray from the centre of the hand controller to the centre of the target in question. This angle was calculated on a 2D plane as seen from above, i.e. the elevation components were discarded for both rays.

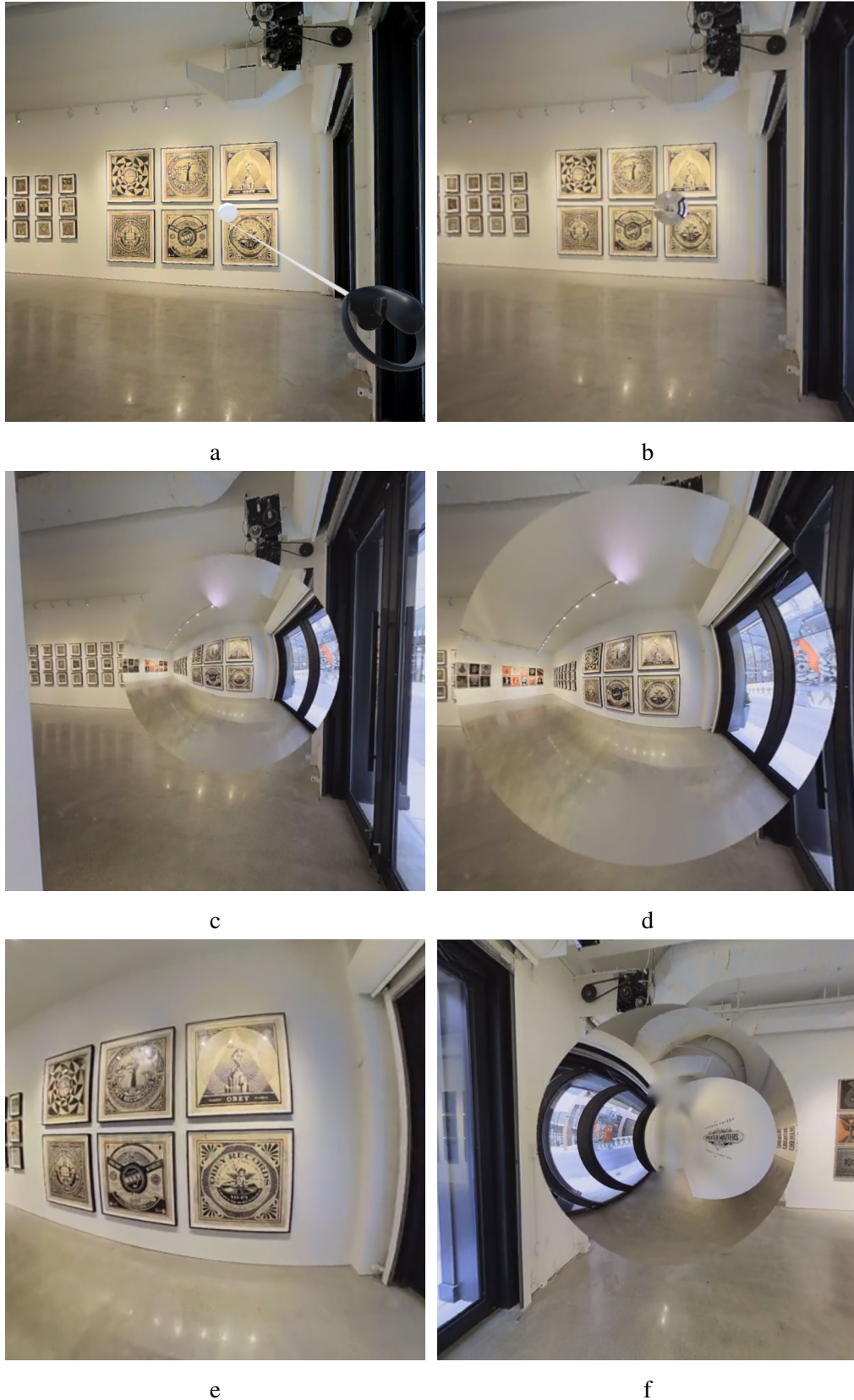
It was expected that the teleportation transition would provide the poorest spatial awareness, while the model transition would provide the best. As the Möbius transition provides some movement cues, it was expected that this transition would produce a spatial awareness result somewhere between the model and the teleportation transitions.

**H1: It was hypothesised that the transition type would have an effect on spatial awareness.**

#### 6.1.3.2 Subjective Measures

**H2: It was hypothesised that the transition type would have an effect on participants’ subjective experience of moving through the space, dizziness, disorientation and naturalness.**

These metrics were assessed by asking the participant to verbally provide a rat-



**Figure 6.3:** Frames from the Möbius transition. Frame a: the user initiates a transition by selecting the new location's marker using the input device. Frames b-c: a zoomed out version of the new location's panorama expands into view. Frames d-e: the Möbius transformation is applied to both panoramas, creating a zooming effect. Frame f: looking backwards, the previous location collapses behind the user.

ing from one to five, where one meant “not at all” and five meant “extremely”. These four questions were, for the transition that they just saw: how much did they feel that they were moving through the space (moving); to what extent did they feel disoriented (disoriented); how dizzy did they feel (dizzy); how natural did the transition feel (naturalness). Naturalness in this context was described to participants as, “how organic and close to real life” the transition felt. Participants were asked to consider each transition on its own. For example, when asking about “disorientation”, participants were told that this was not about their general sense of confusion about the scene, but whether or not they had been disoriented by that transition.

**Moving:** an important characteristic of a transition is whether or not a user feels as if they are moving through the scene. As the teleportation transition is instantaneous, it was expected that users would not feel that they are “moving through the space” during this transition. It was expected that users would strongly feel that they are moving through the space during the 3D model transition. As the Möbius transition provides some movement cues, it was expected that participants would feel somewhat as if they are moving through the space during this transition.

**Dizzy:** in direct contrast to a participant feeling like they are moving through the space, it was expected that teleportation would produce a low rating for dizziness, while Möbius and 3D model transitions would produce higher ratings.

**Disoriented:** as teleportation transitions have previously been shown to disorient users [95], it was expected that this transition would produce a higher subjective rating for disorientation. The model transition was expected to produce the lowest disorientation result, while the effect of the Möbius transition was unclear.

**Natural:** it was expected that the Möbius and teleportation transitions would be rated poorly for naturalness, while the 3D model transition would receive a higher rating.

Additionally, participants’ preferences for transition types were explored. It was unclear what transitions users would like most. This was measured by asking participants to state a binary preference between the last two transitions they saw. A preference value was taken for each possible pairing and ordering of transition types.

**H3: It was hypothesised that there would be a difference in participants’ preferences for transition types.**

### 6.1.3.3 Movement Profile

**H4: It was hypothesised that the transitions type would have an effect on the movement profile of a user.**

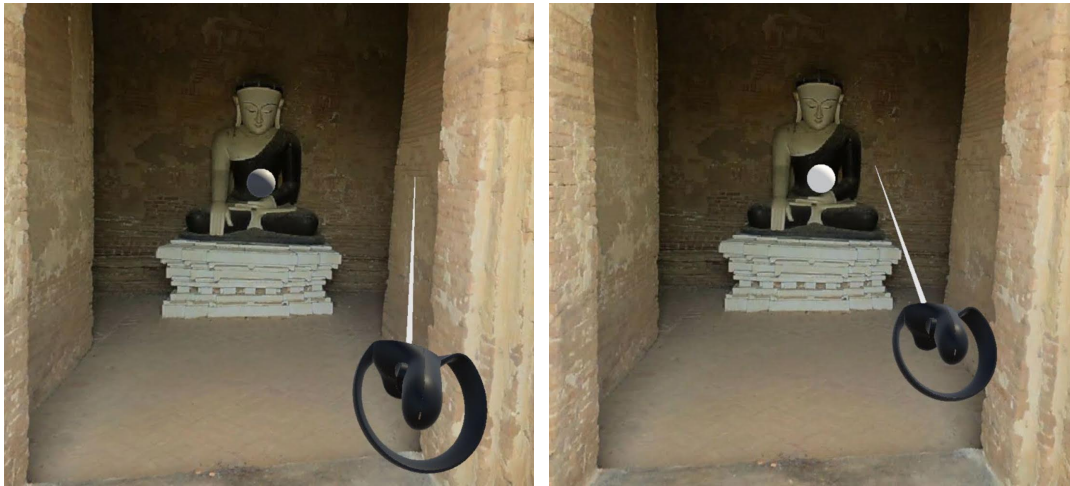
Movement profile was measured by examining the time taken to initiate the next transition, following the completion of the previous transition. As teleportation is more likely to disorient participants, it was expected that it would take longer for a user to initiate the next transition following the teleport transition than following the model or Möbius transitions.

### 6.1.4 Experimental Setup

Participants wore an Oculus CV1 HMD. The CV1 was driven by a Windows 10 desktop PC with an Intel i7-6700 CPU running at 3.4GHz with 32GB of RAM. The video card in use was a NVIDIA GeForce GTX 1080. The software was implemented using the Unity game engine, version 5.6.2f1.

As the captured 360° images must be viewed from the centre of the viewing sphere, only the three degrees of freedom associated with orientation were available to users through their head-tracked movements. This meant there was no visual feedback available to participants in regards to their physical position. As a result of this – coupled with thevection caused by some of the transitions – it was decided that having a participant stand during the study could be unsafe. To ensure safety, participants were seated while wearing the HMD. A swivel chair was used to allow the participant to rotate freely, while the HMD was suspended from the ceiling to avoid any movement limitations that would otherwise have been caused by the cable.

As the experiment required the use of a handheld, tracked input device, an Oculus Touch controller was used. The right-hand controller was used, however as our interface did not require use of the trigger buttons, participants could hold the controller in their preferred hand. Use of the Oculus Touch controller required 360° positional tracking. To facilitate this, three Oculus Sensors were placed in a triangle around the swivel chair facing inwards. This provided accurate positional tracking for the hand controller, as well as drift correction for the rotation tracking of the HMD. When wearing the HMD, participants could see a virtual representation of the hand controller. This virtual representation included a ray, to make it easy for participants to identify



**Figure 6.4:** Floating spheres represent camera locations the user can move to. Spheres are a grey colour when not selected, as shown in the left image. When the user points the hand controller near the sphere, it glows white, as shown in the right image, indicating that when the user presses a button they will be transitioned to that location.

the exact direction the hand controller was pointing. The virtual hand controller maintained the same relative position from the user's head as in the real world, even though the HMD position was not used to update the virtual head position inside the viewing sphere.

Participants were shown four scenes, each of which was captured from multiple locations. Other available locations were represented to the user visually as a sphere, floating at eye height at the location in 3D space that it represented. Pointing the hand controller near the sphere caused the sphere to glow, as shown in Figure 6.4, giving visual feedback to the user that the location was selected. When the user pressed a button on the hand controller, they were transitioned from their current location to their selected location via one of the three transition types described in section 6.1.2. Only other locations in the line-of-sight of the current location were available at any time.

Throughout the study, participants were asked to find targets. Targets were brightly colored squares, placed against walls or columns inside the scene at eye height. An example of such a target is shown in Figure 6.5a. These targets provided a catalyst for exploration, as well as easily identifiable reference points for the pointing task, as discussed later in section 6.1.6.

While a 3D model of the scene was available, it was not photorealistic and contained visual artefacts. The 360° images of the scene were not stereoscopic. To keep

the experience consistent, all visuals were presented monoscopically. This included the spheres that represented other locations, the colored targets, and the 3D model during the model transition. The spheres that represented other locations and the colored targets were only visible when the user was static, and were disabled during transitions.

### **6.1.5 Participants**

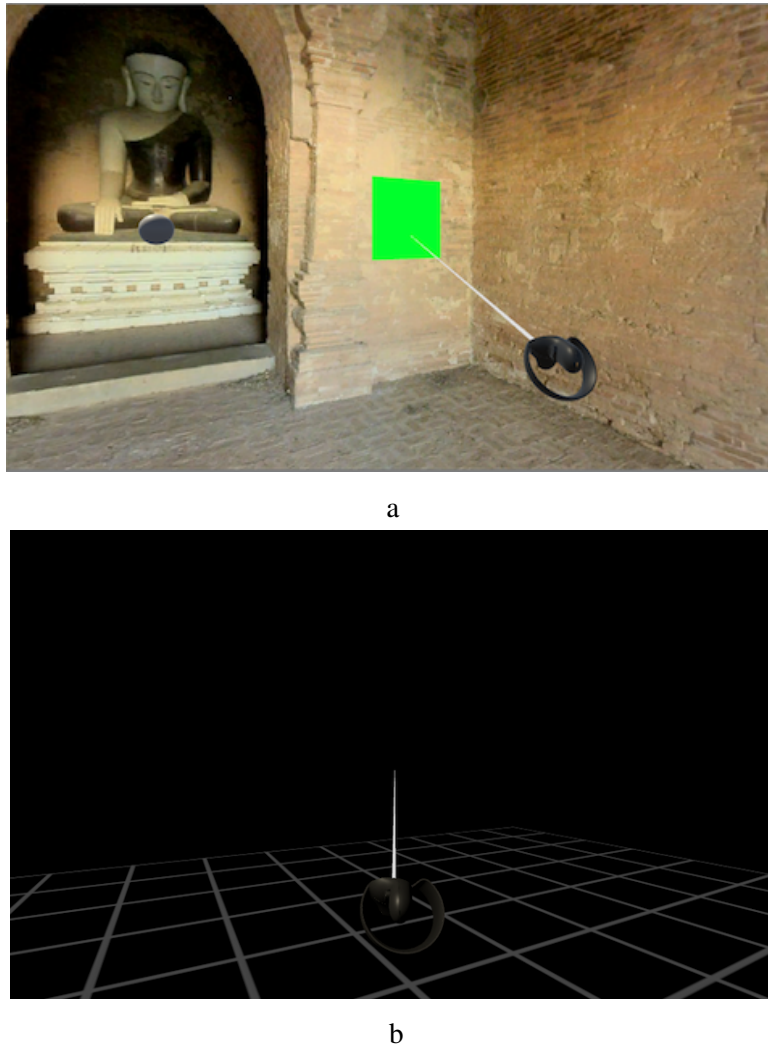
The study was approved by the UCL Research Ethics Committee (project ID 8923/003). All participants were recruited via a participant pool website. Thirty-three participants took part, however data from two were excluded as they did not complete the trial. For one this was due to time limitations, and the second withdrew following discomfort from simulator sickness. Of the remaining 31 participants, 18 were female and 13 male. Ages ranged from 18 to 52 years (mean: 27.68; standard deviation: 7.951). All participants had normal or corrected to normal vision.

### **6.1.6 Experimental Procedure**

Participants were asked to read an information sheet, as well as completing a pre-experiment questionnaire and SSQ. The experimental procedure was then explained to them. As simulator sickness was of particular concern in this experiment due to the large amount of vection involved, the risks of simulator sickness were covered in detail, as was the participant's right to stop at any time.

Participants were asked to sit on a swivel chair, and shown the HMD and the hand controller. Participants then put on the HMD, and were shown a test scene to familiarise them with the equipment and the procedure. A 360° image of a room was shown, and the participant encouraged to rotate in their chair to view the entire room. They were then asked to find a green target, and then hit this target (point at it with the hand controller and press a button). This caused the target to disappear, and another target to appear elsewhere in the room. The participant was then directed to find this new target.

Once the second target had been located and hit, there was a delay of a few seconds before the world faded to a grid environment. In this environment, the entire world is faded out. A grid is faded in at ground level to allow the participant to maintain awareness of their orientation, as shown in Figure 6.5b. The participant was told that when this grid environment appeared, they would be asked to point with the hand controller at where the most recently seen target was in relation to their current location in the



**Figure 6.5:** An example pointing task. A green target is found by the user (frame a). Two transitions of the same type later, the world fades to the grid environment (frame b) and the user points to where they believe that most recently seen target is in relation to their current location in the scene.

scene. At this time, they were asked to point towards the most recently seen target and press a button. When a button was pressed, the grid environment faded out and the virtual scene was faded back in.

Participants were then introduced to the first scene, and instructed on how to transition from one location in the scene to another. They were then asked to move around the space, moving from location to location, looking for and hitting targets.

The experiment proceeded in sets of two targets, called A and B in this example. First, a participant was asked to find target A. Each target was identified by a color, and no participants had any form of colorblindness or experienced any difficulty in identifying targets. An example of such a target can be seen in Figure 6.5a. Once

target A was found, they would be instructed to find target B. To ensure comparability between participants, target A would only be visible from a single location. Although it was not stipulated to the participant, after hitting target A, the available locations were restricted, to ensure the participant had to follow a set route. After moving two locations - in which the participant was shown the same transition type - the world faded to the grid environment after a delay of one second, as shown in Figure 6.5b. The participant was then asked to perform the pointing task, i.e. point at where target A was from their current location and press a button. The error angle was then recorded. After completion of the pointing task, the world was faded back in and the user continued looking for target B. After locating target B, they were instructed to return to target A in as few transitions as possible. Returning to target A – referred to as the *returning phase* – allows for analysis of a natural user movement pattern without interference from the target search i.e. when returning to target A, the user generally knows where they are going, and are therefore not scanning the space for targets.

In total there were nine such pairs of targets across the four scenes, resulting in nine pointing tasks. In order to balance any difference in pointing task difficulty between transition type conditions, the order of transitions shown before each pointing task was counterbalanced. With three transitions under test (3 = 3D model, T = teleportation, M = Möbius), six orderings were possible (3TM, 3MT, T3M, TM3, M3T, MT3). Each participant saw a single ordering of transitions – for example a participant assigned to the first ordering would have performed their nine pointing tasks after seeing transitions in the order 3TM3TM3TM.

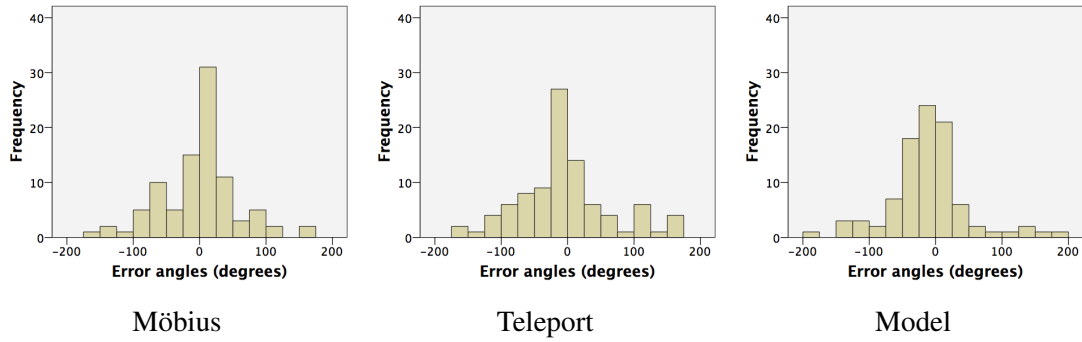
At times the experimenter would ask the participant a question about their subjective experience of the transitions. Five such questions were possible. Four of these questions – moving, dizziness, disorientation and naturalness, as described in section 6.1.3.2 – required the participant to provide a rating from one to five. Before a transition, the experimenter would get the system to select one of these four questions. In order to balance any ordering or possible scene effects, the system selected these questions randomly. Following the transition, if the question had been asked before for that transition type, the question was skipped. If the question had not been asked before, the experimenter would pause the environment (the environment faded out slightly, and actions by the participant were disabled) and orally ask the participant to rank

the transition for that metric. Once the participant had provided their response orally, the experimenter would unpause the environment and the participant would continue searching for the next target. If the experimenter accidentally asked the same question twice for one transition type, the mean of those ratings was used.

The fifth possible question was for the participant to specify which of the last two transitions they preferred. In a pilot study, it became clear that participants were unable to remember the second-to-last transition with enough clarity to provide an accurate comparison. To ensure that the participant was able to provide an answer, they were alerted two transitions in advance that the experimenter was going to ask their preference, allowing them to be actively comparing them. The experimenter also pressed a button, ensuring the system would show two different transitions. The pairs of transitions shown to the user were programmed to ensure that the participant saw each possible pairing of transitions in all orders. With three transition types, this meant the user provided a preference for all six possible pairings.

The experimenter determined when to ask these questions through observation of the participant's position, and attempted to avoid asking questions at any point when the question would interfere with other metrics. For example, if the participant was one location away from a target, the experimenter would not ask the user which of the next two transitions they preferred, as at the next location the participant may have found the target and initiated a pointing task. Likewise, the experimenter avoided asking questions during the returning phase, as this may have affected the participant's movement profile.

Due to the amount of vection involved, participants took a five minute break between each scene. This was done to reduce the cumulative effects of simulator sickness. During this break, participants were asked to complete a pen-and-paper map-placement task. This task was intended to provide a general idea of a participant's spatial awareness of the scene. Participants were asked to mark camera and target locations on a map of the environment they had just seen. Following the final scene, participants were asked to complete an SSQ again. Participants were then debriefed, and given £10 compensation for taking part.



**Figure 6.6:** Histograms of all pointing task results, including the angle's sign, for each transition type.

## 6.2 Results

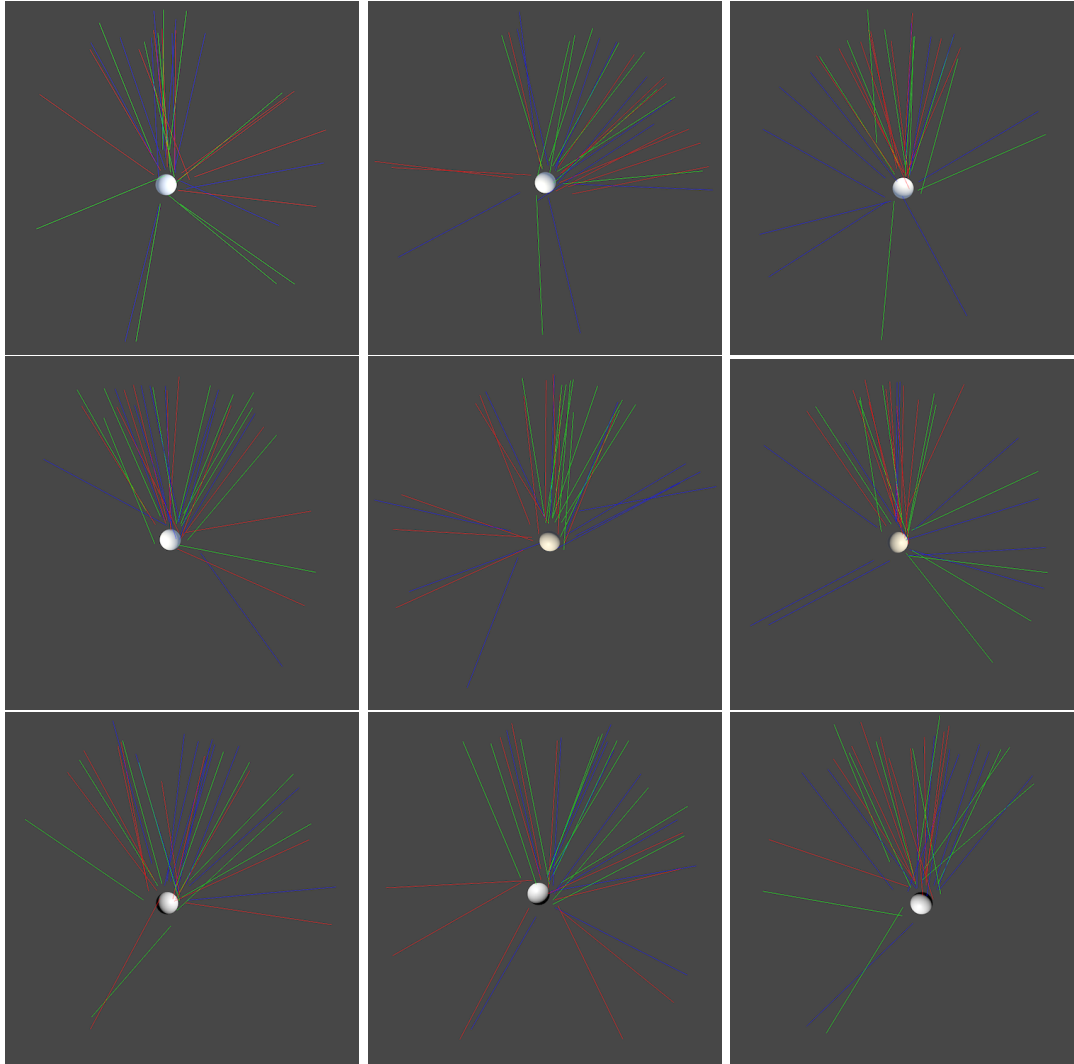
### 6.2.1 Spatial Awareness

#### 6.2.1.1 Error Angle

The error angle was defined as the angle between the user's pointing ray and the ray from the centre of the hand controller to the centre of the target in question, calculated on a 2D plane as seen from above. When all pointing task error angles – including their sign – are shown in a histogram, the distribution shows the expected peak near  $0^\circ$ , as shown in Figure 6.6. Although this data roughly follows a bell shape, analysis with Shapiro-Wilk indicates that the data cannot be considered normally distributed, most likely due to the frequency and spread of extreme data points. Pointing task error angles were often extremely high, indicating that participants struggled with this task. Indeed, participants frequently had error angles over  $90^\circ$ .

The difficulty of a pointing task is related to the two transitions between when the user sees the target, and when they are asked to point at it. This could be affected by the distance travelled in those two transitions, the angle between the two transitions, and the level of scene occlusion involved. For each pointing task, the two transitions taken were predetermined; only one route option was available to the user at these times, so each pointing task is directly comparable to itself between participants. For a visualisation of all pointing task answers, see Figure 6.7. A breakdown of participant error angles by pointing task is shown in Table 6.2.

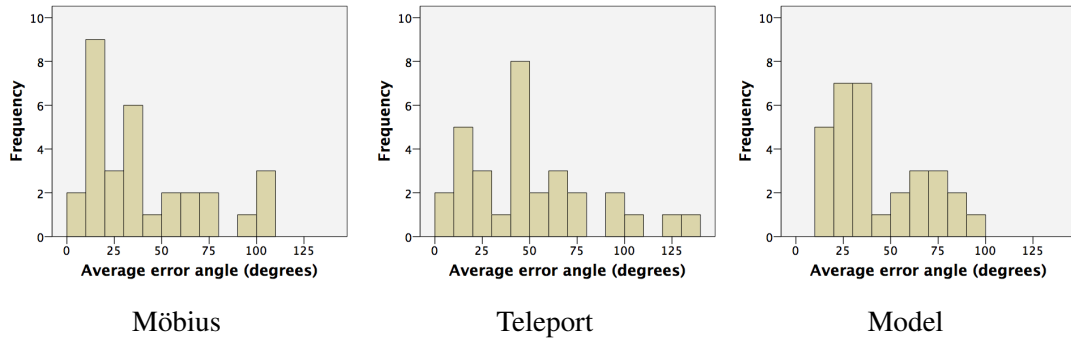
When performing statistical analysis, the average of all three pointing task error angles was taken for each transition type. Only the magnitude of the error angle was



**Figure 6.7:** All participant answers for all pointing tasks. Each coloured line shows the ray from the hand controller at the point the participant pressed a button, indicating they were pointing towards the target. Blue = teleport, red = Möbius, green = model. The sphere in the centre represents the viewing sphere for the pointing task location i.e. the fixed location of the participant's head. Each image is oriented so the vertical aligns with the line between the centre of the viewing sphere and the centre of the target in question i.e. the target the user is trying to point at is due north of the sphere in each image.

**Table 6.2:** Mean and standard deviation for each pointing task.

Pointing task	3D model		Möbius		Teleport		All	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Gallery	63.65	64.33	42.4	32.46	43.79	57.73	49.7	51.85
Upstairs 1	47.09	48.83	61.35	24.04	60.49	52.1	56.01	42.76
Upstairs 2	35.58	52.87	17.69	9.73	71.29	55.63	38.6	45.01
Upstairs 3	33.01	25.81	40.34	38.37	33.66	42.68	35.82	35.34
Temple 1	24.71	34.37	36.04	41.2	66.83	46.02	41.95	43.19
Temple 2	52.14	51.16	12.69	10.94	68.81	44.62	43.06	44.31
Downstairs 1	46.3	36.45	54.58	52.82	25.59	24.18	42.56	40.71
Downstairs 2	35.56	20.07	83.22	58.11	54.49	48.39	57.04	47.43
Downstairs 3	42.34	43.3	20.23	20.73	30.38	37.15	31.65	35.5

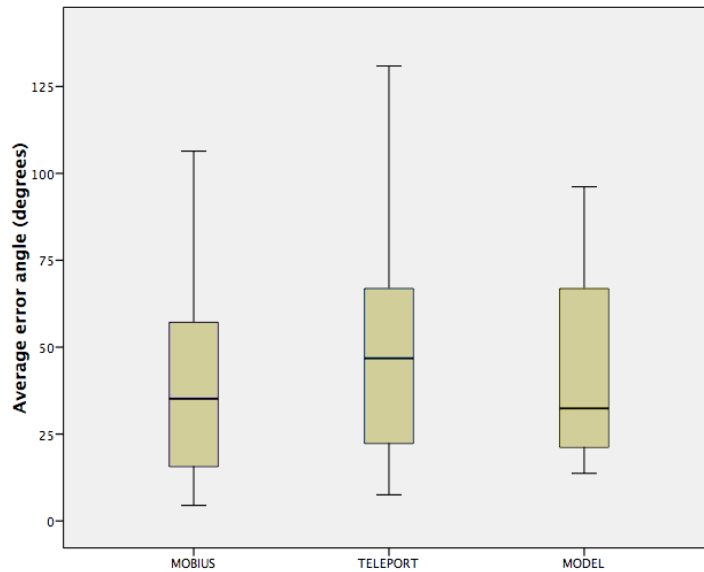
**Figure 6.8:** Histograms of mean pointing task results for each transition type.

used. Large error angles, however, were frequent in the data. When the average of three pointing tasks was taken, this often resulted in unrepresentative values. For example, a participant with error angles  $11.7^\circ$ ,  $13.1^\circ$  and  $115.7^\circ$  following teleportation transitions results in an average of  $46.8^\circ$ . A participant having a large error angle was so frequent in the data that taking the average of three pointing tasks for each transition type resulted in the data appearing multimodal, as shown in Figure 6.8.

As the data is non-normal, a non-parametric test was required. A Friedman test was run to determine if there were differences in absolute pointing task error angles following three different transition types. Error angle increased from model (Mdn =  $32.38^\circ$ ), to Möbius (Mdn =  $35.18^\circ$ ), to teleport (Mdn =  $46.81^\circ$ ), but the differences were not statistically significant,  $\chi^2(2) = 2.774$ ,  $p = .250$ .

### Post Hoc Analysis

It is clear that participants struggled with this task, and the data does not fit expectations. The median values for error angles appear quite different between transition types, with teleport producing a median value 45% larger than the model transition. The high variability and large number of extreme values, however, make analysis difficult. As a result, we examined the data for possible post hoc analysis techniques.



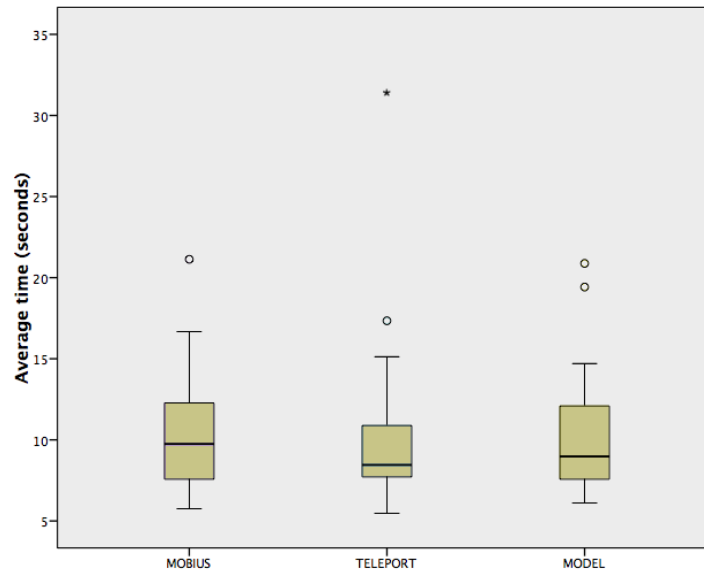
**Figure 6.9:** Boxplot of average pointing task results for each transition type.

The average pointing task data contained no outliers, with an outlier being defined as a data point beyond 1.5 times the interquartile range (IQR). This can be seen in the boxplot shown in Figure 6.9, with whiskers representing 1.5 times the IQR. This is due to the high frequency with which participants produced high error angles, resulting in a large IQR, and means the data is not suitable for filtering out outliers.

An alternative approach is to define a sensible cutoff value, and remove all values above that cutoff across all transition types. The issue with this technique is that – as the teleport transition tended to have more extreme data points – filtering out extremely high error angles across all transition types introduces bias, and results in a reordering of mean values.

Counting the frequency of extreme data points would be one way to identify how often a participant became completely disoriented. An issue with this type of analysis is that the cutoff value is arbitrary, and could be selected such as to force a statistically significant result. As a result, we did not examine the frequency of extreme values.

In previous work that has used a pointing task, it was found that self-reported gaming experience played a role in pointing task performance [99]. In this work, data from gamers and non-gamers was separated, with the data from gamers having lower variance and producing a statistically significant result. A Spearman’s rank-order test, however, found no correlation between frequency of playing video games and pointing task performance in our data. Participant conformity during the pen-and-paper map-



**Figure 6.10:** Boxplot of average times between fading to grid environment and completion of pointing task.

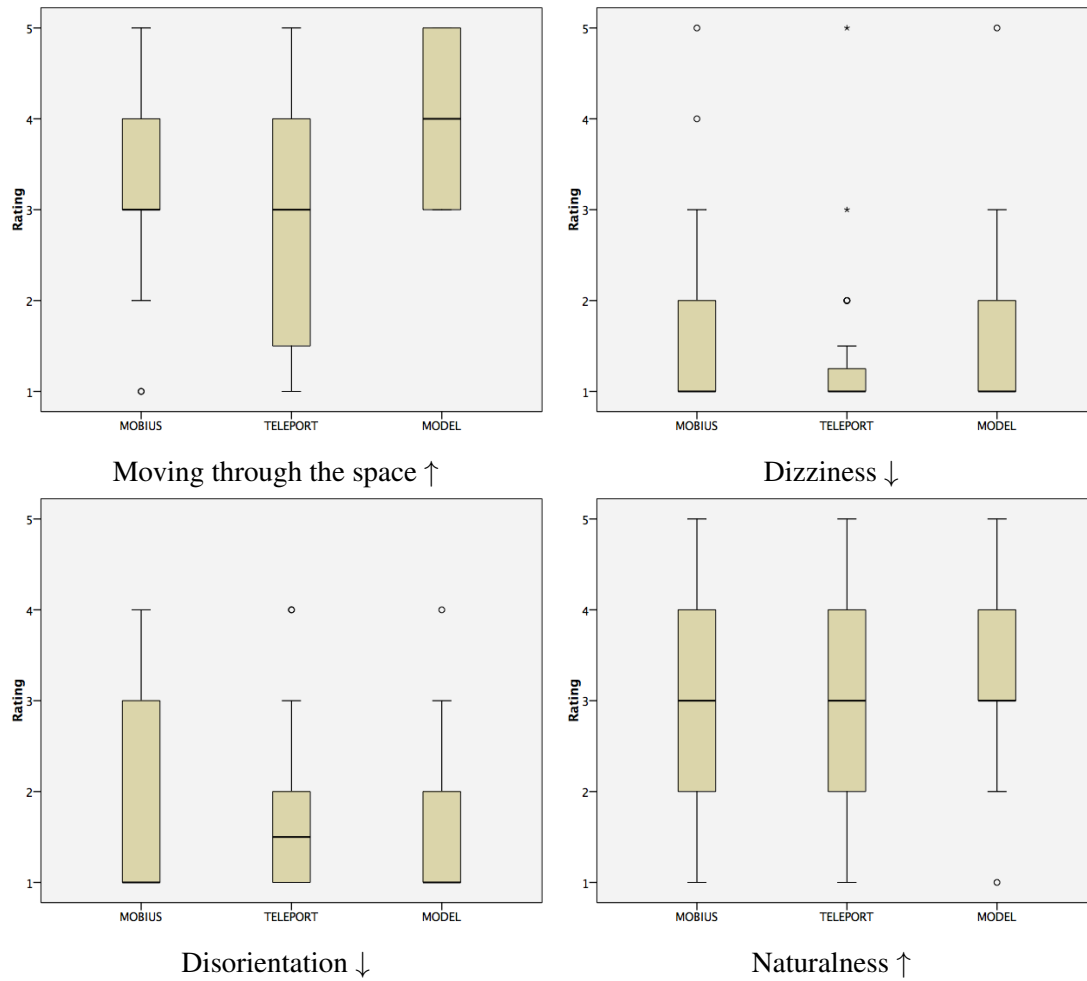
placement task appeared highly variable, so this data was not analysed.

### 6.2.1.2 Time

As can be seen in Figure 6.10, there were outliers in the time taken to complete the pointing task. As a result, a Friedman test was run to determine if there were differences between transition types. The median time taken to complete the pointing task increased from teleport (Mdn = 8.453), to model (Mdn = 8.975), to Möbius (Mdn = 9.756), but the differences were not statistically significant,  $\chi^2(2) = 1.613$ ,  $p = .446$ . There was little difference in the mean time to complete the pointing task between transition types, with teleport averaging 10.0s (SD = 4.8), model averaging 10.1s (SD = 3.7) and Möbius averaging 10.4s (SD = 3.7).

## 6.2.2 Subjective Measures

As subjective measures were given on a five point scale, it could be argued that the data was of an interval type. The data, however, was not normally distributed as assessed by Shapiro-Wilks, and contained outliers. As a result, parametric techniques were not appropriate. Therefore we treated the data as ordinal, and used the non-parametric Friedman test for analysis. Boxplots for all rated subjective metrics are shown in Figure 6.11.



**Figure 6.11:** Boxplots of subjective ratings. ↑ indicates a higher rating is better, ↓ indicates a lower rating is better.

### 6.2.2.1 Moving Through the Space

A Friedman test was run to determine if there were differences in participants' subjective experience of moving through the space for the three different transition types. Participants' ratings for moving through the space were statistically significantly different for different transitions,  $\chi^2(2) = 18.907$ ,  $p < .0005$ . Pairwise comparisons were performed using pairwise Wilcoxon signed-rank tests with a Bonferroni correction for multiple comparisons. Means are included here due to equal median values. Post hoc analysis revealed a statistically significant increase in the subjective experience of moving through the space from teleport (Mdn = 3.0, mean = 2.742) to model (Mdn = 4.0, mean = 3.875) ( $p < .0005$ ) and teleport to Möbius (Mdn = 3.0, mean = 3.419) ( $p = .042$ ), but not between model and Möbius ( $p = .063$ ).

### 6.2.2.2 Dizzy

A Friedman test was run to determine if there were differences in participants' subjective experience of feeling dizzy during the three different transition types. Means are included here due to equal median values. Participants' ratings for feeling dizzy increased from teleport (Mdn = 1.0, mean = 1.389), to model (Mdn = 1.0, mean = 1.625), to Möbius (Mdn = 1.5, mean = 1.719), but the differences were not statistically significant,  $\chi^2(2) = 2.774$ ,  $p = .250$ .

### 6.2.2.3 Disoriented

A Friedman test was run to determine if there were differences in participants' subjective experience of feeling disoriented by the three different transition types. Participants' ratings for feeling disoriented increased from model (Mdn = 1.0), to Möbius (Mdn = 1.25), to teleport (Mdn = 1.5), but the differences were not statistically significant,  $\chi^2(2) = 2.136$ ,  $p = .344$ .

### 6.2.2.4 Naturalness

A Friedman test was run to determine if there were differences in participants' subjective experience of naturalness during the three different transition types. Means are included here due to equal median values. Participants' ratings for naturalness decreased from model (Mdn = 3.0, mean = 3.317), to Möbius (Mdn = 3.0, mean = 2.952), to teleport (Mdn = 3.0, mean = 2.911), but the differences were not statistically significant,  $\chi^2(2) = 2.482$ ,  $p = .289$ .

## 6.2.3 Preference

Binary preference data was analysed by fitting a Bradley-Terry model [129]. Here, a “contest” is considered to mean when a participant was asked to state a preference between two transitions, with the preferred transition becoming the “winner” of that contest. The count of wins for each transition type are shown in Table 6.3.

The parameters of the Bradley-Terry model were estimated using maximum likelihood. The model aims to estimate the probability that transition type  $i$  would beat transition type  $j$  in a contest, for each possible pairing of transition types  $i$  and  $j$ , where  $i \neq j$ . A positive-valued ability score  $\alpha$  is calculated for each transition type, such that the odds that  $i$  will beat  $j$  are  $\alpha_i/\alpha_j$ .

**Table 6.3:** Transition preferences.

Winner	When playing against		
	Model	Teleport	Möbius
Model	-	35	46
Teleport	25	-	39
Möbius	14	20	-

The model can be expressed in the logit-linear form

$$\text{logit}[pr(i \text{ beats } j)] = \lambda_i - \lambda_j,$$

where  $\lambda_i = \log \alpha_i$  for all  $i$ . This allows all of the parameters  $\{\lambda_i\}$  to be estimated using standard generalised linear models (GLM).

Analysis was conducted using the BradleyTerry2 package for R [130]. As the parameters are relative rather than absolute, the 3D model parameter  $\lambda_{\text{model}}$  is set to zero as an identifying convention.

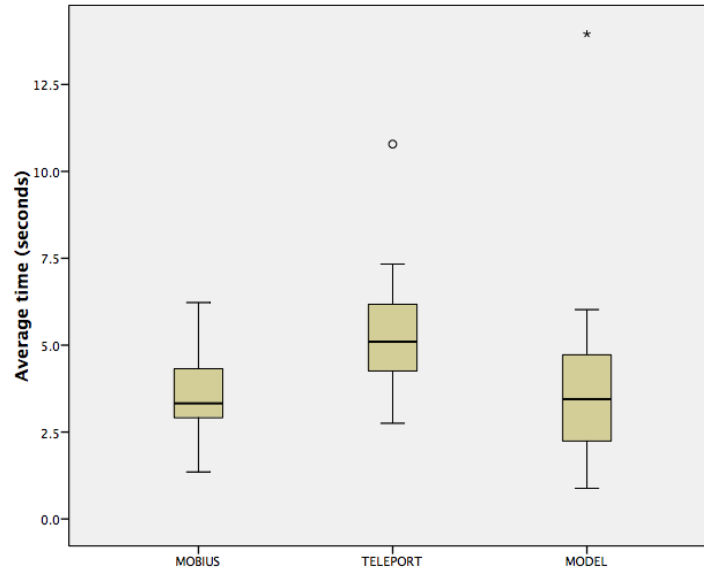
Preference counts for transition types decreased from model (wins = 82,  $\lambda_{\text{model}} = 0$ ) to teleport (wins = 63,  $\lambda_{\text{teleport}} = -0.3904$ ) to Möbius (wins = 34,  $\lambda_{\text{mobius}} = -1.1180$ ).

As the model can be expressed as a GLM, it is possible to calculate an analysis of deviance table and perform a chi-squared likelihood ratio test to obtain significance values.

Participant preferences were statistically significantly different for different transitions (GLM:  $\chi^2(2) = 25.744$ ,  $p < .0005$ ). Post hoc analysis was performed by fitting pairwise Bradley-Terry models, with a Bonferroni correction for multiple comparisons. Post hoc analyses revealed a statistically significant decrease in preference from model (wins = 82) to Möbius (wins = 34) ( $p < .0005$ ) and teleport (wins = 63) to Möbius ( $p = .038$ ), but not between model and teleport ( $p = .59$ ).

### 6.2.4 Movement Profile

In order to explore the movement profile of the user following the different transition types, a time metric was examined following each transition during the returning phase (i.e. when a participant was returning to a known target location, and therefore not searching for a new target). The metric was the time between the completion of one transition and the initiation of the next transition. A boxplot of these results is shown in Figure 6.12.



**Figure 6.12:** Boxplot of average time before initiation of the next transition during the returning phase.

The average time data contained outliers, as assessed by visual inspection of the boxplots. As a result, a Friedman test was run to determine whether there were statistically significant differences in the average time before the next transition following each of the three transition types. The time before initiating the next transition was statistically significantly different for different transition types,  $\chi^2(2) = 21.355$ ,  $p < .0005$ . Pairwise comparisons were performed using pairwise Wilcoxon signed-rank tests with a Bonferroni correction for multiple comparisons. Post hoc analysis revealed a statistically significant increase in the time before the next transition from Möbius (Mdn = 3.33) to teleport (Mdn = 5.1) ( $p < .0015$ ) and from model (Mdn = 3.45) to teleport ( $p < .0015$ ), but not from model to Möbius ( $p = 1$ ).

### 6.2.5 SSQ

The mean Total Severity (TS) score for the pre-experiment SSQ was 4.46 (SD = 7.81), while the mean TS for the post-experiment SSQ was 26.66 (SD = 32.13). TS values were calculated using the formula specified by Kennedy et al. [108]. While this is well below a high mean TS of around 70 [131], the increase from pre-exposure to post-exposure scores does indicate that simulator sickness may be an important factor in this context.

## 6.3 Discussion

### 6.3.1 Spatial Awareness

It is clear participants struggled with the pointing task, as is evidenced by the frequency of extremely high values in the error angles metric. The time taken to complete the pointing task was also high, with the mean for all three transition types being around 10s from fading to the grid environment to the participant pressing the button indicating they were pointing towards the target. This may indicate that MV360M experiences do not promote good spatial awareness. This is an important consideration, as an expected improvement in spatial awareness may be one of the most compelling reasons to employ such systems over single-view 360° media.

The poor spatial awareness results may also have been caused by the implementation of the experiment. As participants sat on a swivel chair, and could therefore not gauge their movement fully through proprioception, a loss of orientation may have been experienced. In two or three instances, participants had pushed themselves round and were rotating freely when the pointing task initiated, so had very poor orientation when the world faded to the grid environment. The grid environment should have ensured that participants retained visual cues about rotation, however, so the frequency of high error rates is perhaps still unexpected.

It is worthwhile to note that, while the procedure followed a pattern of pairs of targets as discussed in section 6.1.6, participants were not told of this pattern and in general did not appear to identify it. Each pointing task seemed to be unanticipated by participants. This means they may not have been making a specific effort at that time to maintain awareness of their own location or the location of the target. This could in part explain why the pointing task results are unexpectedly high.

There is also evidence that humans generally struggle with pointing tasks. In a study by Ruddle et al., a pointing task was used to assess spatial awareness in virtual buildings when viewed in a HMD or a desktop display [132]. In this study, participants navigated through a large-scale virtual environment using a keyboard and mouse in the desktop condition, and a handheld input device in the HMD condition. Based on visual inspection of their boxplots, the pointing task in the HMD condition produced average error angles of around 45°, while the average for the desktop display was around 55°.

The standard error of mean (SEM) was approximately  $5^\circ$  for both. In comparison, our average error angles were between  $41^\circ$  and  $51^\circ$ , with a SEM between  $4.5^\circ$  and  $6^\circ$ . Their study, however, required participants to navigate buildings containing around 70 rooms – a task that would likely be considered substantially harder than ours. Other work in this area reveal similarly poor pointing task results (e.g. [133]). In retrospect, we believe our pointing task was not well designed for measuring spatial awareness. We would encourage future studies to either make the task less difficult, for example by not fading to a grid environment, or by choosing a different method for measuring spatial awareness.

### 6.3.2 Subjective Ratings

#### 6.3.2.1 Disorientation

The subjective ratings for naturalness, disorientation and dizziness showed very little variation between transition types, with no differences nearing statistical significance. Each subjective question was asked once per transition type per participant. The question was randomly generated by the system, meaning the experimenter could not influence for which locations the question was asked. It was clear, however, that following occasional teleport transitions, some participants became disoriented. This usually presented itself through a verbal indicator from the participant. Anecdotally, this appeared to be more common when a teleportation ended close to a wall, meaning the participant had few visual features with which to orient themselves. Occasionally, participants would teleport twice in quick succession, resulting in confusion. The measurement method in use was not able to capture these events. The metric does indicate, however, that the transition type did not generally have an impact on the subjective experience of disorientation.

#### 6.3.2.2 Dizziness

The dizziness metric also did not establish any significant differences between transition types. Similarly to disorientation, there were occasional transitions where participants indicated verbally that they had felt dizzy, but our measurement method was not able to capture these. Anecdotally, these were during model transitions in which the locations were unusually far apart, meaning the user was moved faster to cover the larger distance over the 5s transition. Additionally, looking around during transitions

seemed to increase dizziness.

While a simulator sickness questionnaire was administered before and after the experience, as all participants experienced all transitions roughly equally, it would not be possible to assess the cumulative simulator sickness effects of any individual transition type from our data.

#### 6.3.2.3 Naturalness

During the Möbius transition, participants often commented on how it felt “weird”. It is perhaps surprising, then, that the naturalness metric did not detect any significant differences between transition types.

#### 6.3.2.4 Moving Through the Space

It is clear from these results that the model and Möbius transitions created a stronger feeling of moving through the space than teleport transitions. This is not surprising for the model transition – as the user does virtually move through the space – but it is an interesting finding for the Möbius transition. The Möbius transition is image-based, and no additional information such as parallax is introduced. That the transition can induce the feeling of moving through the space means it could be a useful tool to allow the easy production of MV360M content that elicits this feeling, without the expense or complexity of reconstructing a 3D model of the scene.

#### 6.3.2.5 Preference

As backed up by the quantitative preference results, the model and teleport transitions were generally well received. Participants expressed different opinions, with some preferring the teleport transition because it was faster, and some preferring the model transition because it was more fun or provided additional information about the scene. In general, participants tended not to enjoy the Möbius transition. Several participants reported finding it “weird”. One participant commented that it felt “like being pulled through a keyhole into a different space”. As stated earlier, however, our implementation is just one possible way to use the Möbius transformation for transitions. Other, more visually pleasing transitions may improve the user preference results. For example, in their work on the topic, Segerman et al. included the visual device of a picture frame to provide a join between two panoramas [124].

### 6.3.3 Movement Profile

As the teleport transition is instantaneous, it is perhaps not a surprise that following a teleportation the user takes more time to initiate the next transition than for the other two transition types. While we initially proposed that a longer delay before the next transition may indicate disorientation, during the study several other factors presented themselves as possible contributors to this effect. During the Möbius and model transitions, the user may have more time and visual information to decide on their next transition – saving them time on arrival. Additionally, the Möbius and model transitions give the user time to adjust their orientation during travel, allowing them to be facing in their desired direction on completion of the transition.

As the model and Möbius transitions each take 6 seconds to complete (0.5 seconds blend in, 5 seconds in transit, and 0.5 seconds blend out), teleporting would still be faster despite the increase in time before the next transition. It is interesting to note, however, that the Möbius and model transitions may not add as much total time to a journey as expected, as users take approximately 1.6 seconds longer on average following a teleport transition to initiate the next transition.

It is also interesting to note that there was very little difference between the Möbius and model transitions in terms of the delay before initiating the next transition. While the Möbius transition could feasibly have disoriented users, causing an increased delay, this does not appear to have been the case. Indeed, the median delay for the Möbius transition is slightly smaller than that of the model transition.

## 6.4 Limitations and Future Work

Although the SSQ was used to evaluate the simulator sickness effects of the experience on participants, it is not possible to identify which transitions contributed most to simulator sickness from our study design. Such an investigation would be valuable, as simulator sickness is likely to play an important role in the adoption of certain transition types.

As transition types were generally randomised, participants may have been unable to fully acclimatise to one transition type. Allowing a participant to acclimatise may be important to investigate aspects such as spatial awareness, as these may be affected by learning. Additionally, as participants acclimatise to the virtual space, their preference

may change, with the faster teleport transition potentially becoming preferred over the more informative model transition.

In our MV360M content, the user explored scenes in which cameras were arranged in a connected network, with each camera being in the line-of-sight of at least one other camera. This allowed the entire scene to be explored, with available locations being visualised to the user by way of a sphere at that location's position in the virtual space. This may not be the case for all MV360M content, as some scenes may be too sparsely captured for such a network to be feasible. Our research does not cover such scenarios, and the transition types explored may not be easily adaptable to these types of content.

Due to our desire to explore a 3D model transition, the available content was limited to static scenes. While our results may be applicable to dynamic scenes, there are other issues that were not addressed. As a result, further work is needed in order to understand the impact of dynamic MV360M on users.

There is a wide variety of transition types to be explored, including variations of the three transition types discussed here. For example, the Möbius and model transitions took six seconds each, irrespective of distance travelled. Varying the time based on the distance travelled could be one way to provide further information to the user. Additionally, our implementation of the Möbius transition is only one way to incorporate the Möbius transformation into a transition, and more complex transformations could potentially improve the visual appearance of image-based transitions. While there are many options, the methods in this paper highlight some important considerations, and show which metrics may be most sensitive to the transition type.

## 6.5 Conclusion

Our research investigates the impact of different transition types in MV360M for static scenes, in which users can navigate around a captured virtual space via a connected network of panoramic views. The three transition types explored were teleportation, a linear move through a 3D model of the scene, and an image-based Möbius transformation. The metrics investigated were spatial awareness, users' movement profiles, transition preference and the subjective feelings of moving through the space, disorientation, dizziness, and naturalness.

Our results indicate that the transition type has a significant impact on the subjec-

tive feeling of moving through the space, with the 3D model and Möbius transitions producing a stronger feeling of moving through the space than the teleport transition. The transition type also had a significant effect on a user's movement profile, with users taking on average 1.6 seconds longer to initiate the next transition following a teleport transition than a 3D model or Möbius transition. The subjective feelings of naturalness, disorientation and dizziness were not significantly different between transition types. A pointing task was unable to identify any significant difference in spatial awareness between transition types. These results indicate that the choice of transition type may have an impact on several aspects of the user's experience when exploring MV360M, and as a result content creators must think carefully before selecting a transition type.

## **Chapter 7**

# **Conclusion**

In the work presented here, we explored the acquisition of 360° media (Chapter 3), and aimed to improve the production pipeline for 360° media experiences (Chapter 4). We also aimed to provide a deeper understanding of the end-user experience of single-view 360° video (Chapter 5) and multi-view 360° media (Chapter 6). This closing chapter summarises the work presented in this thesis. We then conclude with a broader look at the field of captured virtual reality, and where we believe future work may take us.

## **7.1 Project summaries**

### **7.1.1 360° media acquisition**

In Chapter 3, the process by which 360° media was acquired for all stages of the project was discussed. This included the capture of content in collaboration with BBC R&D and Middlesex University. These captured pieces included a test documentary, shot using a Ladybug3 camera, and a longer documentary piece about a beekeeper shot using a GoPro HERO camera array. Additionally, the resourcing of single-view 360° media from YouTube and multi-view 360° media from Matterport were also covered. A complete list of all 360° media used throughout this work is available in Appendix B.

### **7.1.2 Object removal in 360° media**

In Chapter 4 we explored object removal in 360° images. A method was proposed in which field-of-view expansion using retargeting techniques was combined with Graph-cut Textures to remove objects near the equator of the viewing sphere. Several extensions and refinements were proposed to improve this technique, including how it can be

extended to remove objects anywhere on the viewing sphere. The proposed retargeting techniques – seam carving and non-homogeneous warping – helped alleviate the distortions this method of object removal can introduce. In some situations, however, such as the presence of very regular surrounding structure, it was shown that this method may fail to produce acceptable results.

Chapter 4 also examined inpainting as a method to remove objects in 360° images, and explored how the choice of projection affects the inpainting result. In many cases where the object to be removed is not at the poles, it was shown that the standard equirectangular projection can be used without additional steps. Due to the distortion characteristics of equirectangular images, however, it was shown that more work needs to be done to inpaint content at the poles. Tripod removal was shown to work well by adding rotation steps to the process, while cubic and rectilinear projections are required for complex inpainting tasks such as geometric textures. Finally, our inpainting techniques were shown to work for video in certain situations, such as tripod removal for a static camera. Inpainting techniques similar to ours are now available in popular post-production tools, such as the 360° video editing software SkyBox Studio [134], which has been acquired by Adobe and integrated into Adobe CC 2018 [135].

### 7.1.3 Cinematic virtual reality

In Chapter 5, we investigated the advantages and disadvantages of CVR compared to traditional viewing formats such as TV. We explored the consumption of panoramic videos in three different display systems: a HMD, a standard 16:9 TV, and a SV+ display. The SV+ display was designed to mitigate some of the issues of directorial control associated with HMD playback, while attempting to retain some of the immersive characteristics.

To investigate the impact of display type on the viewing experience, an experiment was conducted that measured spatial awareness, narrative engagement, enjoyment, memory, fear, guiding attention, and a viewer's concern about missing something. Our results indicated that the HMD offered a significant benefit in terms of enjoyment and spatial awareness, and our SV+ display offered a significant improvement in enjoyment over traditional TV for one of the videos. We were unable to confirm the work of a previous study that showed incidental memory may be lower in a HMD over

a TV.

Drawing attention and a viewer's concern about missing something were also not significantly different between display conditions. These were unexpected results, and may indicate that using a fixed chair in a HMD experience places a soft limit on the field of regard. This finding may have important consequences for video production, as at-home viewers are likely to be seated on a couch.

#### **7.1.4 Multi-view 360° media**

In Chapter 6, we investigated the effect of transition types on immersive MV360M experiences. A repeated-measures experiment was conducted with 31 participants. Wearing a head-mounted display, participants explored four static scenes using three transition types: teleport, a linear move through a 3D model of the scene, and an image-based transition using a Möbius transformation. The metrics investigated were spatial awareness, users' movement profiles, transition preference and the subjective feelings of moving through the space, disorientation, dizziness, and naturalness.

Our results indicated that there was no significant difference between transition types in terms of spatial awareness. In retrospect, however, we believe our choice of metric for spatial awareness may have been inappropriate. Significant differences were found for users' movement profiles, with participants taking 1.6 seconds longer to select their next location following a teleport transition. While this might indicate that the teleport transition disoriented viewers, we also proposed several other factors that likely played a role in this, such as the additional time available to participants during the model and Möbius transitions to adjust their orientation during transit.

The model and Möbius transitions were found to be significantly better in terms of creating the feeling of moving through the space than the teleport transition. This is an interesting finding for the Möbius transition. The Möbius transition is image-based, and no parallax information is introduced. As a result, this transition could be a useful tool to allow the creation of MV360M experiences that elicit the feeling of moving through the space, without the expense of reconstructing a 3D model of the scene.

Preference was also significantly different, with model and teleport transitions being preferred over Möbius transitions. Our implementation, however, is just one possible way to use the Möbius transformation for transitions. Other, more visually

pleasing transitions may be able to improve the user preference results.

It is clear from our results that trade-offs between transitions will require content creators to think carefully about what aspects they consider to be most important when producing MV360M experiences.

## 7.2 Future work

This work aimed to improve the production pipeline for 360° media, and establish methods for the evaluation of such experiences. Due to the commercial value in improving the production pipeline, there has been substantial investment in this space from companies such as Adobe [136] and The Foundry [137]. As these popular tools now support inpainting, the discussion of how best to apply inpainting techniques described in Chapter 4 is more pertinent than ever, as content creators can now easily implement these techniques in their media. Our FOV expansion and Graph-cut Textures technique represents an alternative method of removing objects from scenes where inpainting will not produce plausible results, and future work could include implementing this technique for popular software packages.

Evaluation of 360° media experiences continues to present challenges. In this work, a range of media types and genres were explored using an array of metrics. Despite this breadth and depth of analysis, these investigations only lay the foundation for the development of a 360° media evaluation framework. There are many further avenues of research to be explored. Some are logical follow-on studies that are clearly indicated by our findings as being of value, and some are entirely novel ways of addressing the issues.

Our findings in Chapter 5 indicated that the user experience was impacted not simply by the display in use, but also by the physical set-up of the experience, and in particular by chair type. The effect of the chair type is something that merits further investigation. An MSc project in this area, co-supervised by the author, was completed at UCL [138]. The thesis presents some interesting findings. The project found evidence that more restrictive chairs discourage exploration and lead to reduced spatial awareness over chairs that fully rotate when viewing 360° videos. An understanding of the impact of such factors is critical for the design of these experiences, and further work in this area would help illuminate some of these issues.

The evaluation of transitions in MV360M presented in Chapter 6 also presents many avenues for development. Exploring dynamic, video-based content would present many additional complexities both to the user, and in terms of evaluation. There are also different interfaces and scenarios to be considered. For example, experiences in which the transitions are not initiated by the user, but are predetermined by a director. Evaluation methods for these experiences would likely need to consider cognitive load, as understanding may become a critical factor as complexity increases.

A major issue that our work was unable to address is the transportive nature of 360° media. While there is an assumption that 360° media creates some form of presence, in which the viewer feels they are physically in the portrayed space and perhaps involved in the action, measuring such a response presents many challenges. While measuring presence remains a challenge for the VR community at large, it is perhaps particularly difficult in 360° media experiences, where a lack of agency may disrupt proxies often considered to indicate presence, such as response-as-if-real [74]. Identification of a suitable proxy for evaluating presence in 360° media experiences would be of substantial benefit to the community.

### **7.3 The future of the field**

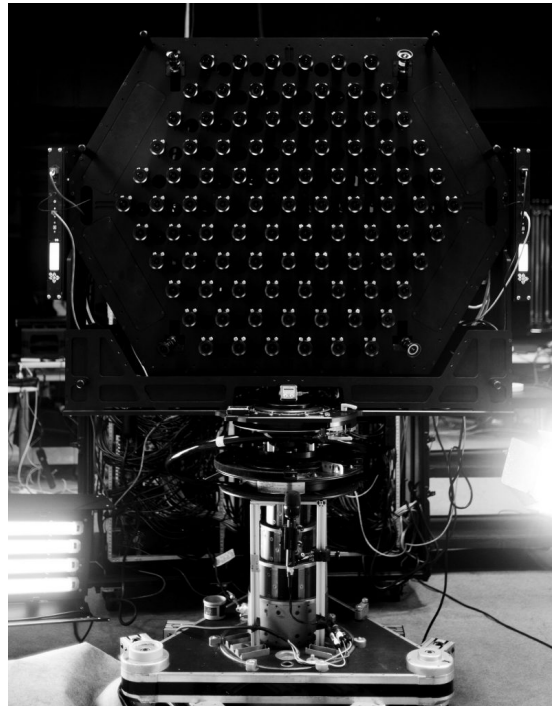
It is clear that 360° media will continue to improve in many areas. Even over the four year duration of this project, the technologies have advanced. These advances cover all areas of the workflow, from production, through distribution, to consumption. Capture has been made easier with purpose built cameras. An example of such a camera is the Z-Cam S1, which allows the comparatively easy creation of 6k monoscopic 360° video at 30fps [139]. Companies like Google continue to push the boundaries of what high-end cameras are capable of. In collaboration with Yi, Google have developed the Yi Halo, capable of filming 8k stereoscopic 360° video at 30fps [140]. Post-production has been simplified, with direct integration of 360° support in software such as Adobe's Premiere Pro and After Effects [134]. Distribution and consumption has also been streamlined, with native support from YouTube. YouTube also supports immersive viewing in Google Cardboard and Daydream HMDs. This increase in quality brings an equally large increase in storage and transmission requirements, however the relevant technologies for these issues also continue to improve. While these developments will

incrementally make 360° media more appealing, they are unlikely to invalidate our findings as they do not alter the fundamentals of the experience.

There is a growing body of data and research exploring 360° media. YouTube now provide heatmaps of where the viewport was directed during 360° media playback, as well as statistics on average 360° media viewing habits. One such statistic is that “people spent 75% of their time within the front 90 degrees of a video” [141]. It is statistics such as this that recently led YouTube to encourage creators to consider using a HFOV of only 180°, a technique termed by YouTube as VR180 [142]. While VR180 may improve the resolution, decrease the required bandwidth and simplify the creation process over 360° media, more research is required to establish the impact that this decreased field-of-view has on the end-user experience. Additionally, it is important to consider that YouTube does not have any data on the physical set-up of the viewer, including what chair type they are using, and they do not provide details on if videos were being watched on an immersive display.

There is also a growing body of academic research looking at CVR. Bindman et al. recently explored the impact of the level of immersion of the viewing device. They examined the impact of immersion on feelings of presence, narrative engagement and empathy when watching 360° media [143]. In their work, 65 participants watched a 360° video on either a HMD or a smartphone. In line with our results, the display device was not found to have an impact on narrative engagement. Additionally, their results did not indicate that the display type had an impact on empathy, but a higher level of immersion did produce stronger levels of self-reported presence.

One area that will require active work to address is the lack of visual grammar available to 360° filmmakers. The visual grammar used in film today was built up over decades. For example, cross-cutting is frequently used to create suspense [144]. In this editing technique, frequent cuts are made between two events, to establish to the audience that both are happening concurrently. Such techniques are not necessarily transferrable to 360° video, and as a result a new visual language must be created. It is this lack of visual grammar that has prompted creators to compare the current state of the field to the early days of cinema (e.g. [145]). While the development of a visual grammar will likely be a critical aspect for the capabilities of storytelling in the field, this is a challenge for creators, and is unlikely to be resolved through academically-led



**Figure 7.1:** Lytro’s Immerse light-field camera. Each circle is an individual capture camera. Image source: Lytro, Inc.

technical solutions.

The technical future of the field is open to debate. One opinion is that the next major step in captured virtual reality will be the introduction of motion parallax. Allowing users to move their head with six degrees of freedom, while maintaining photorealism, is an immense technical challenge. There are currently two suggested approaches to creating “volumetric video”, i.e. video that supports motion parallax when a viewer in a HMD moves their head. While the terminology for these technologies is evolving, here we refer to them as “light-field capture” and “free-viewpoint video”.

Light-field capture is a technology in which multiple cameras are used to capture a “light-field”, a space in which perspective-correct views are available from anywhere inside that volume [146]. Light-field video capture systems have existed for some time, but have generally been restricted to capturing in a specific viewing direction [147]. Even the most up-to-date light-field camera from Lytro, a leader in the field, only captures in a single direction, and must be rotated in order to capture a full 360° scene [148]. An image of this camera is shown in Figure 7.1. To our knowledge, there are currently no omnidirectional light-field capture systems.

In free-viewpoint video capture, a large number of cameras are arranged facing



**Figure 7.2:** Microsoft's free-viewpoint video capture system [120].

inwards towards a capture volume [26]. While older versions of such systems were limited to capturing character performances, more recent systems are more generic and can capture any central object [120]. The central object can then be segmented from the background, and the various views integrated to create a textured 3D model. For VR viewing, this 3D model could then be integrated into scenery that is computer generated or has been captured using photogrammetry. Such systems have the benefit of allowing greater freedom of movement to viewers, who can view the captured scene from any angle. Despite continuing work in the field, however, these systems are limited in terms of their visual quality, with resolution and artefacts that mean these experiences do not currently reach photorealism [120]. An example of a free-viewpoint video system is shown in Figure 7.2.

Whatever the future of captured virtual reality experiences may be, it is clear that evaluation of such experiences is a critical issue that presents many challenges. There is a complex interplay of perceptual and cognitive processes at work. Regardless of where the technology may take us, a concrete understanding of these processes is essential to ensuring that these experiences are engaging and safe, and it is hoped that the work presented here can act as a foundation for such evaluations.

## **Appendix A**

# **List of Acronyms**

**ANNF** Approximate nearest-neighbour field

**AR** Augmented reality

**CVR** Cinematic virtual reality

**DoF** Degrees of freedom

**FOR** Field of regard

**FOV** Field of view

**HFOV** Horizontal field of view

**HMD** Head-mounted display

**IEQ** Immersive Experience Questionnaire

**IMU** Inertial measurement unit

**MNEQ** Measuring Narrative Engagement Questionnaire

**MV360M** Multi-view 360° media

**NNF** Nearest-neighbour field

**SA** Spatial awareness

**SAR** Spatial augmented reality

**SSQ** Simulator Sickness Questionnaire

**VFOV** Vertical field of view

**VE** Virtual environment

**VR** Virtual reality

## Appendix B

# List of all 360° media

Included here is a list of all 360° media used throughout this document. The media is split into two categories: images and video that are viewable only from a single position at any time (“360° images and video”) and media that is viewable from multiple positions at any time (“multi-view 360° media”).

## 360° images and video

---

Name: Sherbrooke Forest  
Source: Peter Gawthrop  
Used: Chapter 4  
URL: <https://www.flickr.com/photos/gawthrop/3241996032>

---



---

Name: Test documentary  
Source: Filmed in collaboration with BBC R&D  
Used: Frames from video used in Chapter 4  
URL: not available online

---



---

Name: Test footage  
Source: Peter Boyd Maclean  
Used: Frames from video used in Chapter 4  
URL: not available online

---



---

Name: London documentary  
Source: BBC R&D  
Used: Chapter 4  
URL: not available online

---



Name: BBC office tour  
 Source: BBC R&D  
 Used: Frames from video used in Chapter 4  
 URL: not available online



Name: The Resistance of Honey  
 Source: Filmed in collaboration with BBC R&D  
 Used: Chapter 5 (documentary stimulus)  
 URL: <https://www.youtube.com/watch?v=t6u3opMTCV4> (to reduce length, edited out 2m20-2m46 and 3m26-6m15)



Name: Noa Neal 'Graffiti'  
 Source: YouTube  
 Used: Chapter 5 (music video stimulus)  
 URL: <https://www.youtube.com/watch?v=LByJ9Q6Lddo>



Name: Hide and Seek  
 Source: YouTube  
 Used: Chapter 5 (horror stimulus)  
 URL: <https://www.youtube.com/watch?v=ePf7mQJ3IvE>



Name: Real Memories  
 Source: YouTube  
 Used: Chapter 5 (narrative stimulus)  
 URL: <https://www.youtube.com/watch?v=9Ta2Et4jX4Y>



## Multi-view 360° media

---

Name: Bagan Buddha Temple

Source: Matterport

Used: Chapter 6 (temple stimulus)

URL:

<https://matterport.com/3d-space/bagan-four-buddha-temple/>

---



Name: House

Source: Matterport

Used: Chapter 6 (upstairs/downstairs stimulus)

URL: not available online

---



Name: Gallery

Source: Matterport

Used: Chapter 6 (gallery stimulus)

URL: not available online

---



## Appendix C

# User study one: questions and tasks for all metrics

### C.1 Pre-experiment questionnaire

Gender (options: male/female)

Age (free text entry)

How many times have you used a head-mounted display in the past? (options: Never/Once or twice/A few times/Moderately often/Often)

On average, how many hours a week do you think you spend watching TV or movies? (options: Never, Rarely, Occasionally, Often, Always)

Do you easily become deeply involved in movies or TV dramas? (options: Never, Rarely, Occasionally, Often, Always)

Do you ever become so involved in a television program or book that people have problems getting your attention? (options: Never, Rarely, Occasionally, Often, Always)

Do you ever become so involved in doing something that you lose all track of time? (options: Never, Rarely, Occasionally, Often, Always)

Do you ever become so involved in a movie that you are not aware of things happening around you? (options: Never, Rarely, Occasionally, Often, Always)

How frequently do you find yourself closely identifying with the characters in a story line? (options: Never, Rarely, Occasionally, Often, Always)

Do you ever become so involved in a daydream that you are not aware of things happening around you? (options: Never, Rarely, Occasionally, Often, Always)

How good are you at blocking out external distractions when you are involved in some-

thing? (options: Very bad/Moderately bad/Neither good nor bad/Moderately good/Very good)

How afraid of bees would you say you are? (options: Not at all/A little/A moderate amount/A fairly large amount/Extremely)

## C.2 DOCUMENTARY stimulus

### Memory

All memory questions were answered using free text entry. To ensure fairness, the answers were marked blind, i.e., the display condition of the participant was not known when their score was tallied.

**Question:** How many years has the beekeeper been recording bees?

**Corresponding audio:** “People have been recording bees since the 1950s – mainly to try and predict swarming. I started recording bees about ten years ago.”

**Question:** How many microphones were on the honeycomb?

**Corresponding audio:** “On this frame there’s a couple of microphones. I like to record stereo inside the hive.”

**Question:** What is the beekeeper allergic to?

**Corresponding audio:** “I’m allergic to any animals with fur: cats, dogs, sheep, horses. So beekeeping seemed the right thing for me.”

**Question:** When did people start recording bees?

**Corresponding audio:** “People have been recording bees since the 1950s – mainly to try and predict swarming.”

**Question:** What are the three types of bees in the hive?

**Corresponding audio:** “Inside each bee hive there are three types of bees: the queen, the workers and the drones.”; “In a beehive, there one queen, anything up to 50 thousand workers, and about 500 or 1000 drones.”; “This baby bee synth emulates the frequencies of bees. It has three oscillators, one for the queen, one for the workers, and one for the drones.”

**Question:** Originally, what was the main reason for people to start recording bees?

**Corresponding audio:** “People have been recording bees since the 1950s – mainly to try and predict swarming.”

**Question:** What ratio of people are allergic to bee stings?

**Corresponding audio:** “I’m not allergic to strings, which one in every thousand people are.”

**Question:** What is the life span of a worker bee?

**Corresponding audio:** “The worker bee gives itself tirelessly to the colony. It’s got a short lifespan, only six weeks.”

**Question:** What percentage of honey is water?

**Corresponding audio:** “Honey has 17% water, which makes it a great conductor, or resistor.”

**Question:** What percentage of animals on the planet did the bees tell the beekeeper are wild rather than domestic?

**Corresponding audio:** “They tell me they’re worried about the world. How there’s only, like, 5% or less wild animals in the world. The other 95% are all domesticated animals.”

## Enjoyment

Enjoyment questions were answered on a 5-point Likert scale of agreement. Considering the display and the video separately, please indicate how much you agree with each statement:

I enjoyed watching this video.

I enjoyed using this display.

## Concern about missing something

Concern about missing something questions were answered on a 5-point Likert scale of agreement.

At times, I was worried I was missing something.

My concern about missing something impacted my enjoyment of the video.

## C.3 HORROR stimulus

### Fear

Fear questions were answered on a 5-point Likert scale of agreement.

I felt nervous while watching this video.

I felt afraid while watching this video.

### Concern about missing something

Concern about missing something questions were answered on a 5-point Likert scale of agreement.

At times, I was worried I was missing something.

My concern about missing something impacted my enjoyment of the video.

### Enjoyment

Enjoyment questions were answered on a 5-point Likert scale of agreement. Considering the display and the video separately, please indicate how much you agree with each statement:

I enjoyed watching this video.

I enjoyed using this display.

### Attention

Who initiated the phone call? (choose from images of:)

The murderer

This victim

A character who doesn't appear on screen

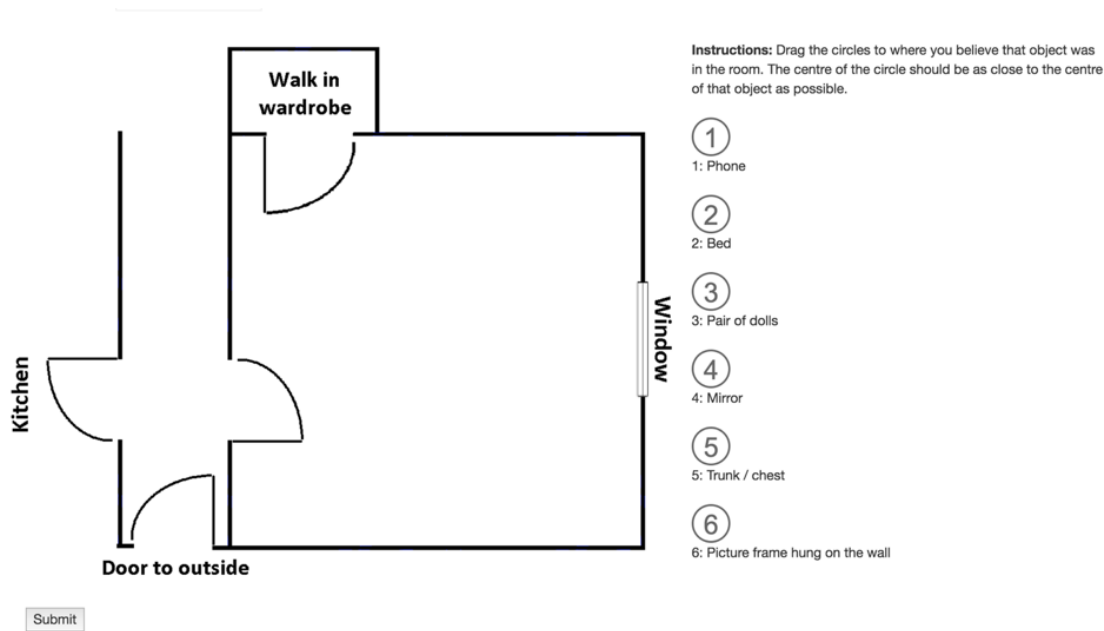
I'm not sure

### Narrative engagement

The questionnaire was the measuring narrative engagement questionnaire (MNEQ). See [86] for full list of questions.

### Spatial awareness

The spatial awareness task was completed on a laptop, using a touchpad to drag and drop circles onto a map of the scene. The task is shown in Figure C.1.



**Figure C.1:** Object placement task for HORROR stimulus.

## C.4 NARRATIVE stimulus

### Attention

The below attention questions were answered with free text entry. To ensure fairness, the answers were marked blind, i.e., the display condition of the participant was not known when their score was tallied.

In the first car scene, what was the younger man holding?

In the ‘flashback’, what was the murder weapon? Describe it as specifically as you can (object type, material, colour).

What was the projection showing? Be as specific as you can.

### Enjoyment

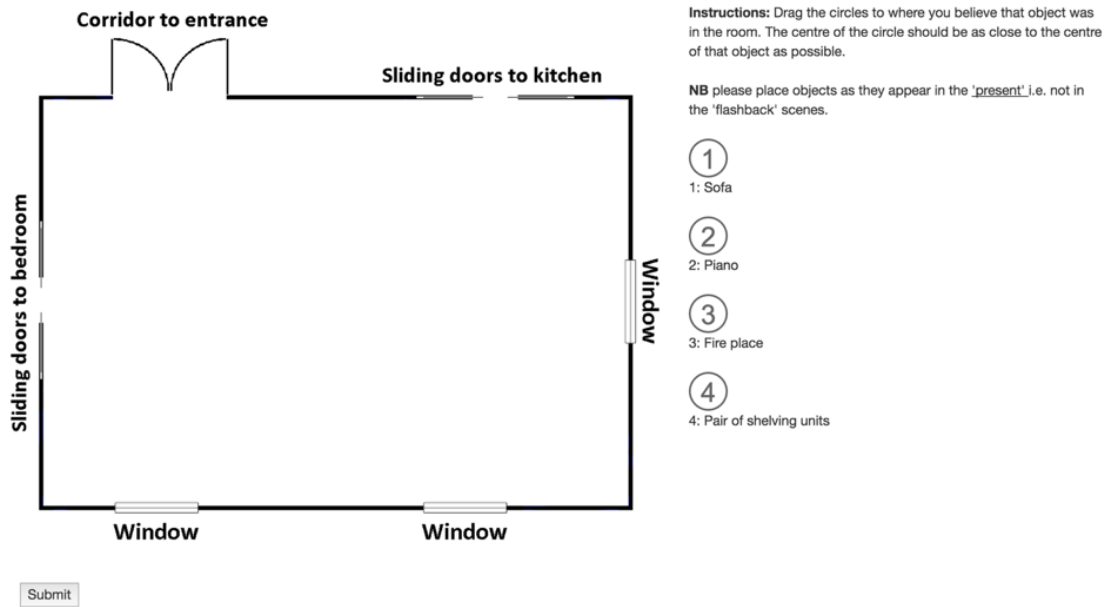
Enjoyment questions were answered on a 5-point Likert scale of agreement. Considering the display and the video separately, please indicate how much you agree with each statement:

I enjoyed watching this video.

I enjoyed using this display.

### Concern about missing something

Concern about missing something questions were answered on a 5-point Likert scale of agreement.



**Figure C.2:** Object placement task for NARRATIVE stimulus.

At times, I was worried I was missing something.

My concern about missing something impacted my enjoyment of the video.

### **Narrative engagement**

The questionnaire was the measuring narrative engagement questionnaire (MNEQ).

See [86] for full list of questions.

### **Spatial awareness**

The spatial awareness task was completed on a laptop, using a touchpad to drag and drop circles onto a map of the scene. The task is shown in Figure C.2.

## **C.5 Final questions**

Simulator sickness was measured using the Simulator Sickness Questionnaire (SSQ).

See [108] for full list of questions.

# Bibliography

- [1] BBC R&D. Surround video. <http://www.bbc.co.uk/rd/publications/whitepaper208>. Accessed: 23-March-2018.
- [2] MIT Media Lab. What is object-based media, anyway? <http://obm.media.mit.edu/>. Accessed: 23-March-2018.
- [3] Philips. Ambilight. <http://www.philips.co.uk/c-m-so/televisions/#pillar=ov-pillar-ambilight>. Accessed: 23-March-2018.
- [4] Brett R Jones, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. IllumiRoom: peripheral projected illusions for interactive experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 869–878. ACM, 2013.
- [5] Ramesh Raskar, Greg Welch, Kok-Lim Low, and Deepak Bandyopadhyay. *Shader lamps: Animating real objects with image-based illumination*. Springer, 2001.
- [6] Oliver Grau. *Virtual Art: from illusion to immersion*. MIT press, 2003.
- [7] Richard D. Altick. The panorama in Leicester Square. In *The Shows of London*, chapter 10, pages 128–140. Belknap Press of Harvard University Press, Cambridge, Mass., 1978.
- [8] John Hannavy. *Encyclopedia of nineteenth-century photography*, pages 365–366. Routledge, 2013.
- [9] Matthew D Smith. The specter of cholera in nineteenth-century Cincinnati. *Ohio Valley History*, 16(2):21–40, 2016.

- [10] Steve Mann and Rosalind W Picard. Virtual bellows: Constructing high quality stills from video. In *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, volume 1, pages 363–367. IEEE, 1994.
- [11] Facebook. Introducing Facebook 360 for Gear VR. <https://newsroom.fb.com/news/2017/03/introducing-facebook-360-for-gear-vr/>. Accessed: 04-January-2018.
- [12] Gorillaz. Saturnz barz (spirit house) 360. <https://www.youtube.com/watch?v=lVaBvyzuypw>. Accessed: 25-October-2017.
- [13] Brian Crecente. 5m Gear VR headsets in homes, Samsung confirms. <https://www.polygon.com/2017/1/4/14172210/gear-vr-headsets-sales>. Accessed: 01-November-2017.
- [14] Shenchang Eric Chen. Quicktime VR: An image-based approach to virtual environment navigation. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 29–38. ACM, 1995.
- [15] Dragomir Anguelov, Carole Dulong, Daniel Filip, Christian Frueh, Stéphane Lafon, Richard Lyon, Abhijit Ogale, Luc Vincent, and Josh Weaver. Google street view: Capturing the world at street level. *Computer*, 43(6):32–38, 2010.
- [16] Andrew MacQuarrie and Anthony Steed. Object removal in panoramic media. In *Proceedings of the 12th European Conference on Visual Media Production*, pages 2–11. ACM, 2015.
- [17] Andrew MacQuarrie and Anthony Steed. Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video. In *Virtual Reality (VR), 2017 IEEE*, pages 45–54. IEEE, 2017.
- [18] Andrew MacQuarrie and Anthony Steed. The effect of transition type in multi-view 360° media. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1564–1573, 2018.

- [19] Delight VR. Complete list of VR camera systems 2018. <https://delight-vr.com/blog/complete-list-of-vr-cameras-2017/>. Accessed: 23-March-2018.
- [20] Josh Lowensohn. YouTube now supports 360-degree videos. <http://www.theverge.com/2015/3/13/8203173/youtube-now-supports-360-degree-videos>. Accessed: 23-March-2018.
- [21] Richard Szeliski and Sing Bing Kang. Direct methods for visual scene reconstruction. In *Representation of Visual Scenes, 1995.(In Conjunction with ICCV'95), Proceedings IEEE Workshop on*, pages 26–33. IEEE, 1995.
- [22] Richard Szeliski and Heung-Yeung Shum. Creating full view panoramic image mosaics and environment maps. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 251–258. ACM Press/Addison-Wesley Publishing Co., 1997.
- [23] Matthew Brown and David G Lowe. Automatic panoramic image stitching using invariant features. *International journal of computer vision*, 74(1):59–73, 2007.
- [24] Harpreet S Sawhney and Rakesh Kumar. True multi-image alignment and its application to mosaicing and lens distortion correction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(3):235–243, 1999.
- [25] Nozon. What is presenz? <http://www.nozon.com/presenz>. Accessed: 23-March-2018.
- [26] Joel Carranza, Christian Theobalt, Marcus A Magnor, and Hans-Peter Seidel. Free-viewpoint video of human actors. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 569–577. ACM, 2003.
- [27] nVidia. Capture, stitch, and stream VR content in real-time with VRWorks 360 video SDK. <https://blogs.nvidia.com/blog/2016/07/25/360-degree-video-stitching/>. Accessed: 08-September-2017.

- [28] BigLook360. Use multiple camera views & 360 degree video to enhance live event broadcasts. <http://biglook360.com/2011/11/360-degree-video-3/>. Accessed: 08-September-2017.
- [29] Pitchfork. Watch Sigur Rós perform Kveikur songs live via interactive webcast. <https://pitchfork.com/news/51227-watch-sigur-ros-perform-kveikur-songs-live-via-interactive-webcast/>. Accessed: 08-September-2017.
- [30] John P Snyder. *Flattening the earth: two thousand years of map projections*. University of Chicago Press, 1997.
- [31] John Parr Snyder. *Map projections – A working manual*. Number 1395. USGPO, 1987.
- [32] Robert Carroll, Maneesh Agrawala, and Aseem Agarwala. Optimizing content-preserving projections for wide-angle images. *ACM Transactions on Graphics (TOG)*, 28(3):43, 2009.
- [33] Vivek Kwatra, Arno Schödl, Irfan Essa, Greg Turk, and Aaron Bobick. Graphcut textures: image and video synthesis using graph cuts. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 277–286. ACM, 2003.
- [34] Robert Keys. Cubic convolution interpolation for digital image processing. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29(6):1153–1160, 1981.
- [35] Shai Avidan and Ariel Shamir. Seam carving for content-aware image resizing. In *ACM Transactions on Graphics (TOG)*, volume 26, page 10. ACM, 2007.
- [36] Michael Rubinstein, Ariel Shamir, and Shai Avidan. Improved seam carving for video retargeting. In *ACM Transactions on Graphics (TOG)*, volume 27, page 16. ACM, 2008.
- [37] Matthias Grundmann, Vivek Kwatra, Mei Han, and Irfan Essa. Discontinuous seam-carving for video retargeting. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 569–576. IEEE, 2010.

- [38] Lior Wolf, Moshe Guttman, and Daniel Cohen-Or. Non-homogeneous content-driven video-retargeting. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–6. IEEE, 2007.
- [39] Yael Pritch, Eitam Kav-Venaki, and Shmuel Peleg. Shift-map image editing. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 151–158. IEEE, 2009.
- [40] Homan Igehy and Lucas Pereira. Image replacement through texture synthesis. In *Image Processing, 1997. Proceedings., International Conference on*, volume 3, pages 186–189. IEEE, 1997.
- [41] Antonio Criminisi, Patrick Perez, and Kentaro Toyama. Object removal by exemplar-based inpainting. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–721. IEEE, 2003.
- [42] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan Goldman. Patch-match: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (TOG)*, 28(3):24, 2009.
- [43] Adobe. Content-aware fill. <http://adobe.com/technology/projects/content-aware-fill.html>. Accessed: 23-March-2018.
- [44] Yonatan Wexler, Eli Shechtman, and Michal Irani. Space-time video completion. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–120. IEEE, 2004.
- [45] Leonardo K Sacht, Paulo C Carvalho, Luiz Velho, and Marcelo Gattass. Face and straight line detection in equirectangular images. In *Workshop de Visão Computacional. Presidente Prudente, SP, Brasil: FTC-UNESP*, pages 101–106, 2010.
- [46] Zhe Zhu, Ralph R Martin, and Shi-Min Hu. Panorama completion for street views. *Computational Visual Media*, 1(1):49–57, 2015.

- [47] Oculus. Binocular vision, stereoscopic imaging and depth cues. [https://developer.oculus.com/design/latest/concepts/bp\\_app\\_imaging/](https://developer.oculus.com/design/latest/concepts/bp_app_imaging/). Accessed: 27-November-2017.
- [48] Samsung. Gear vr with controller. <http://www.samsung.com/global/galaxy/gear-vr/>. Accessed: 01-November-2017.
- [49] Dominic Brennan. Samsung Gear VR install base has passed 5 million headsets. <https://www.roadtovr.com/samsung-sold-5-million-gear-vr-headsets/>. Accessed: 21-November-2017.
- [50] Greg Welch and Eric Foxlin. Motion tracking: No silver bullet, but a respectable arsenal. *IEEE Computer graphics and Applications*, 22(6):24–38, 2002.
- [51] Timothy J Buker, Dennis A Vincenzi, and John E Deaton. The effect of apparent latency on simulator sickness while using a see-through helmet-mounted display: Reducing apparent latency with predictive compensation. *Human factors*, 54(2):235–249, 2012.
- [52] John Carmack. Latency mitigation strategies. <http://altdevblog.com/2013/02/22/latency-mitigation-strategies/>. Accessed: 27-November-2017.
- [53] Sean Ong. *Beginning Windows Mixed Reality Programming*. Springer, 2017.
- [54] W Hunt. Virtual reality: The next great graphics revolution. *Keynote Talk High-Performance Graphics*, 2015.
- [55] Ian P Howard and Brian J Rogers. *Binocular vision and stereopsis*. Oxford University Press, USA, 1995.
- [56] Carolina Cruz-Neira, Daniel J Sandin, and Thomas A DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 135–142. ACM, 1993.

- [57] A Weffers-Albu, S de Waele, W Hoogenstraaten, and C Kwisthout. Immersive TV viewing with advanced Ambilight. In *Consumer Electronics (ICCE), 2011 IEEE International Conference on*, pages 753–754. IEEE, 2011.
- [58] Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira. Roomalive: Magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 637–644. ACM, 2014.
- [59] Brett Jones, Projection Mapping Central. The illustrated history of projection mapping. <http://projection-mapping.org/the-history-of-projection-mapping/>. Accessed: 23-March-2018.
- [60] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 179–188. ACM, 1998.
- [61] Memo Akten, Marshmallow Laser Feast. Sony PlayStation3 video store. <http://www.memo.tv/sony-playstation3-video-store/>. Accessed: 23-March-2018.
- [62] Bot & Dolly. Box. <https://vimeo.com/75361102>. Accessed: 23-March-2018.
- [63] Ramesh Raskar, Remo Ziegler, and Thomas Willwacher. Cartoon dioramas in motion. In *ACM SIGGRAPH 2006 Courses*, page 6. ACM, 2006.
- [64] Gordon Wetzstein, Oliver Bimber, et al. Radiometric compensation through inverse light transport. In *Pacific conference on computer graphics and applications*, pages 391–399, 2007.
- [65] Doug A Bowman, Joseph L Gabbard, and Deborah Hix. A survey of usability evaluation in virtual environments: classification and comparison of methods. *PRESENCE*, 11(4):404–424, 2002.

- [66] Pieter Seuntjens, Ingrid Vogels, and Arnold van Keersop. Visual experience of 3D-TV with pixelated ambilight. *Proceedings of PRESENCE*, 2007, 2007.
- [67] ITU-R. *Methodology for the subjective assessment of the quality of television pictures*. International Telecommunication Union, 2003.
- [68] Byron Reeves, Annie Lang, Eun Young Kim, and Deborah Tatar. The effects of screen size and message content on attention and arousal. *Media Psychology*, 1(1):49–67, 1999.
- [69] Simone Schnall, Craig Hedge, and Ruth Weaver. The immersive virtual environment of the digital fulldome: Considerations of relevant psychological processes. *International Journal of Human-Computer Studies*, 70(8):561–575, 2012.
- [70] Diana Fonseca and Martin Kraus. A comparison of head-mounted and hand-held displays for 360 videos with focus on attitude and behavior change. In *Proceedings of the 20th International Academic Mindtrek Conference*, pages 287–296. ACM, 2016.
- [71] Mel Slater and Sylvia Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and virtual environments*, 6(6):603–616, 1997.
- [72] Bob G Witmer and Michael J Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments*, 7(3):225–240, 1998.
- [73] Michael Meehan, Brent Insko, Mary Whitton, and Frederick P Brooks Jr. Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics (TOG)*, 21(3):645–652, 2002.
- [74] Mel Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.

- [75] Maria V Sanchez-Vives and Mel Slater. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4):332–339, 2005.
- [76] Mel Slater. How colorful was your day? why questionnaires cannot assess presence in virtual environments. *PRESENCE*, 13(4):484–493, 2004.
- [77] Mel Slater, Andrea Brogni, and Anthony Steed. Physiological responses to breaks in presence: A pilot study. In *PRESENCE 2003: The 6th Annual International Workshop on Presence*, volume 157. Citeseer, 2003.
- [78] Wutthigrai Boonsuk, Stephen Gilbert, and Jonathan Kelly. The impact of three interfaces for 360-degree video on spatial cognition. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2579–2588. ACM, 2012.
- [79] Doug A Bowman, Ameya Datey, Young Sam Ryu, Umer Farooq, and Omar Vasnaik. Empirical comparison of human behavior and performance with different display devices for virtual environments. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 46, pages 2134–2138. SAGE Publications, 2002.
- [80] Elizabeth Louise Glisky. Incidental memory. In *Encyclopedia of Clinical Neuropsychology*, pages 1303–1304. Springer, 2011.
- [81] Richard C Atkinson and Richard M Shiffrin. Human memory: A proposed system and its control processes. *Psychology of learning and motivation*, 2:89–195, 1968.
- [82] David Wechsler et al. *Wechsler memory scale (WMS-III)*. Psychological Corporation San Antonio, TX, 1997.
- [83] Eric D Ragan, Ajith Sowndararajan, Regis Kopper, and Doug A Bowman. The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments. *Presence: Teleoperators and Virtual Environments*, 19(6):527–543, 2010.

- [84] A Rizzo, L Pryor, R Matheis, M Schultheis, K Ghahremani, and A Sey. Memory assessment using graphics-based and panoramic video virtual environments. In *Proc. 5th Intl Conf. Disability, Virtual Reality & Assoc. Tech*, 2004.
- [85] Katerina Mania and Alan Chalmers. The effects of levels of immersion on memory and presence in virtual environments: A reality centered approach. *CyberPsychology & Behavior*, 4(2):247–264, 2001.
- [86] Rick Busselle and Helena Bilandzic. Measuring narrative engagement. *Media Psychology*, 12(4):321–347, 2009.
- [87] Freya Sukalla, Helena Bilandzic, Paul D Bolls, and Rick W Busselle. Embodiment of narrative engagement. *Journal of Media Psychology*, 2015.
- [88] Charlene Jennett, Anna L Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. Measuring and defining the experience of immersion in games. *International journal of human-computer studies*, 66(9):641–661, 2008.
- [89] Mihaly Csikszentmihalyi. *Flow: The psychology of optimal performance*, 1990.
- [90] Evren Bozgeyikli, Andrew Raij, Srinivas Katkoori, and Rajiv Dubey. Point & teleport locomotion technique for virtual reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, pages 205–216. ACM, 2016.
- [91] Kay M Stanney, Ronald R Mourant, and Robert S Kennedy. Human factors issues in virtual environments: A review of the literature. *Presence: Teleoperators and Virtual Environments*, 7(4):327–351, 1998.
- [92] Martin Usoh, Kevin Arthur, Mary C Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P Brooks Jr. Walking > walking-in-place > flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 359–364. ACM Press/Addison-Wesley Publishing Co., 1999.

- [93] Sharif Razzaque, Zachariah Kohn, and Mary C Whitton. Redirected walking. In *Proceedings of EUROGRAPHICS*, volume 9, pages 105–106. Manchester, UK, 2001.
- [94] Sebastian Freitag, Dominik Rausch, and Torsten Kuhlen. Reorientation in virtual environments using interactive portals. In *3D User Interfaces (3DUI), 2014 IEEE Symposium on*, pages 119–122. IEEE, 2014.
- [95] Doug A Bowman, David Koller, and Larry F Hodges. Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques. In *Virtual Reality Annual International Symposium, 1997., IEEE 1997*, pages 45–52. IEEE, 1997.
- [96] Perry W Thorndyke and Barbara Hayes-Roth. Differences in spatial knowledge acquired from maps and navigation. *Cognitive psychology*, 14(4):560–589, 1982.
- [97] Atsuyuki Okabe, Ken Aoki, and Wataru Hamamoto. Distance and direction judgment in a large-scale natural environment: Effects of a slope and winding trail. *Environment and Behavior*, 18(6):755–772, 1986.
- [98] Doug A Bowman, Elizabeth T Davis, Larry F Hodges, and Albert N Badre. Maintaining spatial orientation during travel in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, 8(6):618–631, 1999.
- [99] Shyam Prathish Sargunam, Kasra Rahimi Moghadam, Mohamed Suhail, and Eric D Ragan. Guided head rotation and amplified head rotation: Evaluating semi-natural travel and viewing techniques in virtual reality. In *Virtual Reality (VR), 2017 IEEE*, pages 19–28. IEEE, 2017.
- [100] Rudy P Darken and John L Sibert. A toolset for navigation in virtual environments. In *Proceedings of the 6th annual ACM symposium on User interface software and technology*, pages 157–165. ACM, 1993.
- [101] Leonard McMillan and Gary Bishop. Plenoptic modeling: An image-based rendering system. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 39–46. ACM, 1995.

- [102] Yann Morvan and Carol O’Sullivan. Handling occluders in transitions from panoramic images: A perceptual study. *ACM Transactions on Applied Perception (TAP)*, 6(4):25, 2009.
- [103] Youichi Horry, Ken-Ichi Anjyo, and Kiyoshi Arai. Tour into the picture: using a spidery mesh interface to make animation from a single image. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 225–232. ACM Press/Addison-Wesley Publishing Co., 1997.
- [104] Lawrence J Hettinger, Kevin S Berbaum, Robert S Kennedy, William P Dunlap, and Margaret D Nolan. Vection and simulator sickness. *Military Psychology*, 2(3):171, 1990.
- [105] Kelly S Hale, Kay M Stanney, Behrang Keshavarz, Heiko Hecht, and Ben D Lawson. Visually induced motion sickness: causes, characteristics, and counter-measures. In *Handbook of Virtual Environments: Design, Implementation, and Applications, Second Edition*, pages 647–698. CRC Press, 2014.
- [106] Thomas Brandt, Johannes Dichgans, and E Koenig. Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. *Experimental Brain Research*, 16(5):476–491, 1973.
- [107] David M Johnson. Introduction to and review of simulator sickness research. Technical report, DTIC Document, 2005.
- [108] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.
- [109] YouTube. A new way to see and share your world with 360-degree video. <https://youtube-creators.googleblog.com/2015/03/a-new-way-to-see-and-share-your-world.html>. Accessed: 27-November-2017.

- [110] Maher Saba. Introducing 360 Video on Facebook. <https://newsroom.fb.com/news/2015/09/introducing-360-video-on-facebook/>. Accessed: 29-November-2017.
- [111] Ehsan Sayyad, Pradeep Sen, and Tobias Hollerer. Panotrace: Interactive 3d modeling of surround-view panoramic images in virtual reality. *Proceedings of VRST 17*, 2017.
- [112] Darko Pavić, Volker Schönefeld, and Leif Kobbelt. Interactive image completion with perspective correction. *The Visual Computer*, 22(9-11):671–681, 2006.
- [113] Katy Newton and Karin Soukup. The storyteller’s guide to the virtual reality audience. <https://medium.com/stanford-d-school/the-storyteller-s-guide-to-the-virtual-reality-audience-19e92da57497>. Accessed: 28-March-2018.
- [114] Carolina Cruz-Neira, Daniel J Sandin, Thomas A DeFanti, Robert V Kenyon, and John C Hart. The cave: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–73, 1992.
- [115] VRFocus. Oculus ‘strongly discouraging’ devs from creating jump scares in VR horror. <http://www.vrfocus.com/2016/01/oculus-strongly-discouraging-devs-from-creating-jump-scares-in-vr-horror/>. Accessed: 28-March-2018.
- [116] Omigamedev. Omiplayer: media player for oculus rift. <https://bitbucket.org/omigamedev/omiplayer>. Accessed: 23-March-2018.
- [117] Peter J Passmore, Maxine Glancy, Adam Philpot, Amelia Roscoe, Andrew Wood, and Bob Fields. Effects of viewing condition on user experience of panoramic video. 2016.
- [118] Vanus Vachiratamporn, Roberto Legaspi, Koichi Moriyama, and Masayuki Numao. Towards the design of affective survival horror games: An investigation on player affect. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 576–581. IEEE, 2013.

- [119] Olive Jean Dunn. Multiple comparisons using rank sums. *Technometrics*, 6(3):241–252, 1964.
- [120] Alvaro Collet, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, and Steve Sullivan. High-quality streamable free-viewpoint video. *ACM Transactions on Graphics (TOG)*, 34(4):69, 2015.
- [121] Jacob Freiberg and Reese Muntean. Bagan Four Buddha Temple. <https://matterport.com/3d-space/bagan-four-buddha-temple/>. Accessed: 01-December-2017.
- [122] Frederick Bonato, Andrea Bubka, Stephen Palmisano, Danielle Phillip, and Giselle Moreno. Vection change exacerbates simulator sickness in virtual environments. *Presence: Teleoperators and Virtual Environments*, 17(3):283–292, 2008.
- [123] Oculus. Simulator Sickness. [https://developer.oculus.com/design/latest/concepts/bp\\_app\\_simulator\\_sickness/](https://developer.oculus.com/design/latest/concepts/bp_app_simulator_sickness/). Accessed: 04-December-2017.
- [124] eleVR. Spherical video editing effects with möbius transformations. <http://elevr.com/spherical-video-editing-effects-with-mobius-transformations/>. Accessed: 08-September-2017.
- [125] Douglas N Arnold and Jonathan Rogness. Möbius transformations revealed. 2008.
- [126] Andrew MacQuarrie. The effect of transition type in multi-view 360° media. <https://www.youtube.com/watch?v=XwdVenkQeLY>, 2017. Accessed: 20-March-2018.
- [127] Andrew MacQuarrie. A 360° video of the Möbius effect. [https://www.youtube.com/watch?v=xV\\_hai4HUBU](https://www.youtube.com/watch?v=xV_hai4HUBU), 2017. Accessed: 20-March-2018.

- [128] Andrew MacQuarrie. Spherical image editing. [https://github.com/andrewmacquarrie/spherical\\_image\\_editing](https://github.com/andrewmacquarrie/spherical_image_editing), 2017. Accessed: 20-March-2018.
- [129] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.
- [130] David Firth and Heather L Turner. Bradley-terry models in R: the bradleyterry2 package. *Journal of Statistical Software*, 48(9), 2012.
- [131] Jennifer A Ehrlich and Eugenia M Kolasinski. A comparison of sickness symptoms between dropout and finishing participants in virtual environment studies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 42, pages 1466–1470. SAGE Publications Sage CA: Los Angeles, CA, 1998.
- [132] Roy A Ruddle, Stephen J Payne, and Dylan M Jones. Navigating large-scale virtual environments: what differences occur between helmet-mounted and desktop displays? *Presence: Teleoperators and Virtual Environments*, 8(2):157–168, 1999.
- [133] Anthony E Richardson, Daniel R Montello, and Mary Hegarty. Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & cognition*, 27(4):741–750, 1999.
- [134] Jamie Pence. How to remove a camera rig from 360 footage in After Effects. <https://www.mettle.com/how-to-remove-a-camera-rig-from-360-footage-in-after-effects-skybox-studio-v2-jamie-pence/>. Accessed: 20-December-2017.
- [135] Mettle. Adobe CC 2018 is now available! Includes Skybox integration. <https://www.mettle.com/adobe-cc-2018-is-now-available-includes-skybox-integration/>. Accessed: 20-December-2017.

- [136] TechCrunch. Adobe Acquires Mettles Skybox Tools To Expands Its VR Video Portfolio. <https://techcrunch.com/2017/06/21/adobe-acquires-mettles-skybox-tools-to-expands-its-vr-video-portfolio/>. Accessed: 28-March-2018.
- [137] StudioDaily. The Foundry Ships Cara VR Plug-In for Nuke. <http://www.studiodaily.com/2016/07/the-foundry-ships-cara-vr-plug-in-for-nuke/>. Accessed: 28-March-2018.
- [138] Yang Hong. The effect of chair type on users' viewing experience for panoramic video. Master's thesis, University College London, 2017. <https://uclic.ucl.ac.uk/study/current-taught-course/distinction-projects/17>. Accessed: 28-March-2018.
- [139] Z-CAM. Z-CAM S1. <http://www.z-cam.com/360-vr-camera-s1/>. Accessed: 22-December-2017.
- [140] Sean O'Kane and Nick Statt. This is Google Jumps next-generation VR camera rig. <https://www.theverge.com/circuitbreaker/2017/4/24/15405540/yi-technology-halo-360-vr-google-jump-start-camera>. Accessed: 22-December-2017.
- [141] YouTube Creator Blog. Hot and cold: Heatmaps in vr. <https://youtube-creators.googleblog.com/2017/06/hot-and-cold-heatmaps-in-vr.html> [Online; accessed 29-July-2018].
- [142] YouTube Official Blog. The world as you see it with vr180. <https://youtube.googleblog.com/2017/06/the-world-as-you-see-it-with-vr180.html> [Online; accessed 29-July-2018].
- [143] Samantha W Bindman, Lisa M Castaneda, Mike Scanlon, and Anna Cechony. Am i a bunny?: The impact of high and low immersion platforms and viewers' perceptions of role on presence, narrative engagement, and empathy during an animated 360 video. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 457. ACM, 2018.

- [144] Charles Derry. *The suspense thriller: Films in the shadow of Alfred Hitchcock*. McFarland, 2001.
- [145] Andrew Marantz. Studio 360. <https://www.newyorker.com/magazine/2016/04/25/making-movies-with-virtual-reality>. Accessed: 20-December-2017.
- [146] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42. ACM, 1996.
- [147] Bennett Wilburn, Michael Smulski, Hsiao-Heng Keli Lee, and Mark Horowitz. The light field video camera. *Media Processors 2002*, 4674:29–36, 2002.
- [148] Ben Lang. Lytro’s latest VR light-field camera is huge, and hugely improved. <https://www.roadtovr.com/lytro-immmerge-latest-light-field-camera-shows-major-gains-in-capture-quality/>. Accessed: 8-December-2017.