PERCEIVED ACCELERATION IN STEREOSCOPIC ANIMATION

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

GRADUATE PROGRAMME IN INTERDISCIPLINARY STUDIES
YORK UNIVERSITY
TORONTO, ONTARIO

NOVEMBER 2017

Abstract

In stereoscopic media, a sensation of depth is produced through the differences of images presented to the left and the right eyes. These differences are a result of binocular parallax caused by the separation of the cameras used to capture the scene. Creators of stereoscopic media face the challenge of producing compelling depth while restricting the amount of parallax to a comfortable range. Control of camera separation is a key manipulation to control parallax. Sometimes, stereoscopic warping is used in post-production process to selectively increase or decrease depth in certain regions of the image. However, mismatches between camera geometry and natural stereoscopic geometry can theoretically produce nonlinear distortions of perceived space. The relative expansion or compression of the stereoscopic space, in theory, should affect the perceived acceleration of objects moving through that space. This thesis suggests that viewers are tolerant of effects of distortions when perceiving acceleration in a stereoscopic scene.

Dedication

This thesis is dedicated to my beloved parents, Aasia Shamim and Abdul Rahman Laldin,
without whose love and affection none of my successes would be possible

Acknowledgements

I would first like to thank my thesis advisors Dr. Robert Allison of Department of Electric Engineering and Computer Science and Dr. Laurie Wilcox of the Department of Psychology at York University. The door to Prof. Allison's and Wilcox's office was always open whenever I needed help or had a question about my research or writing. They consistently allowed this paper to be my own work, but steered me in the right the direction whenever they thought I needed it. I am also grateful for the 3DFLIC team for cultivating my interest in 3D film technology by giving me the opportunity to work on several productions, an experience that led to two published papers which are included in this thesis as Chapter 2, which was published as “Motion in depth constancy in stereoscopic display” and Chapter 3, published as “The effects of depth warping on perceived acceleration in stereoscopic animation”.

I would also like to thank other members of the team including Karim Benzeroual, who was very generous in sharing his knowledge of 3D media with me throughout my Masters study. I would like to thank Carly Hylton for her fastidious work in collecting data for the experiments conducted during this study. I would like to particularly thank the Department of Electrical Engineering and Computer Science tech staff, notably Ulya Yigit, Seela Balkissoon, and Paul Griffith for their support in setting up the technical infrastructure for my research. In addition, I would like to thank Fiona Fernandes for her support in navigating the graduate studies program and the Interdisciplinary Department for giving me the opportunity to meld my creative interest and technical interest in the form of this interdisciplinary research.

I would like to thank the funders, namely Ontario Media Development Corporation and Ontario Centres of Excellence, whose grants to the 3DFlic project made my research possible. I would like to extend special thanks to Rob Burton and Arc Productions for expert advice and assistance with the preparation of the stimuli in the second experiment of this thesis.

Finally, I must express my very profound gratitude to my parents, Abdul Rahman Laldin and Aasia Shamim Abdul Rahman, for providing me with unfailing support and continuous encouragement throughout my years of study. I would like to thank my brother Ali for his words of wisdom including "things are getting better one step at a time", my sister Ayesha, for her unwavering support and encouragement, in keeping me on track, and finally my brother, Omar, for providing me with much needed direction and focus on completing my thesis. This accomplishment would not have been possible without all of them supporting me every step of the way. Thank you.

TABLE OF CONTENTS

Abstract ii

Dedication iii

Acknowledgements iv

List of Figures x

List of Variables xi

1 Perception of Stereoscopic 3D Images 1

1.1 Stereoscopic Depth Perception in Humans 2

1.1.1 Size constancy 8

1.1.2 Depth constancy 8

1.1.3 Interaction of depth and size constancy 11

1.2 Stereoscopic Parameters and Process 12

1.2.1 Interaxial Distance 13

1.2.2 Convergence or Zero-Parallax Setting (ZPS) 15

1.2.3 Focal Length 17

1.3 Aspects of projection that impact stereopsis 18

1.3.1 Screen Parallax 18

1.3.2 Screen Size 22

1.3.3 Viewing Distance 24

1.4 Perceptual Distortions 24
1.4.1 Puppet Theatre Effect 25
1.4.2 Cardboard effect 26
1.4.3 Roundness 27
1.5 Non-linear Depth Scaling in S3D 29
1.6 Motion in Depth 32
1.6.1 Target Vergence, Relative Disparity, and Looming 32
1.6.2 Ground Plane Information 36
1.6.3 Perception of Acceleration in Motion in Depth 38
1.7 Summary 40
2 Motion in Depth Constancy in Stereoscopic Displays 42
2.1 Chapter Summary 42
2.2 Introduction 43
2.2.1 Depth Distortions in Stereoscopic Space 45
2.2.2 Motion in Depth Cues in the Visual System 47
2.3 Experiment 1 - Effects of Interaxial on Perceived Acceleration 49
2.3.1 Method 50
2.3.1.1 Subjects 50
2.3.1.2 Apparatus and Stimuli 50
2.3.1.3 Procedure 52

2.3.2 Results 53
2.4 Experiment 2 – Monocular and binocular contributions to acceleration bias .. 56
2.4.1 Method 56
2.4.1.1 Subjects and Apparatus 56
2.4.1.2 Procedure 57
2.4.2 Results 57
2.5 Discussion 59
2.6 Conclusions 65
3 The Effects of Depth Warping on Perceived Acceleration in Stereoscopic Animation 66
3.1 Chapter Summary 66
3.2 Introduction 67
3.3 Methods 72
3.3.1 Participants 72
3.3.2 Stimuli 72
3.3.3 Non-linear parallax re-mapping 76
3.3.4 Stereoscopic Display and Viewing 79
3.3.5 Procedure 79
3.4 Results and Discussion 81
3.5 Discussion 84

4 General Discussion..91

Bibliography 95

List of Figures

Figure 1-1 5

Figure 1-2 7

Figure 1-3 10

Figure 1-4 13

Figure 1-5 16

Figure 1-6 19

Figure 1-7 20

Figure 1-8 29

Figure 2-1 47

Figure 2-2 51

Figure 2-3 51

Figure 2-4 55

Figure 2-5 58

Figure 3-1 73

Figure 3-2 74

Figure 3-3 78

Figure 3-4 81

Figure 3-5 82

List of Variables

V = Viewing distance

D = Relative Distance between fixation point and object

IO = interocular distance

IA = interaxial distance

C = Convergence distance

Z= real depth

Z' = Perceived depth

z_{in}= distance from the camera in the original stereo pairs

z_{out}= the geometrically predicted distance from the viewer of the display

M= size of the screen

W = width of convergence planes

S_p = screen parallax of object P

F= fixation point

f = focal length

a = distance at which the camera is focused

P=point in the scene

1 Perception of Stereoscopic 3D Images

Although there has been an explosion of research and development of Stereoscopic 3D (S3D) cinema in the last few years, there was already an extensive amount of literature on the geometry of stereoscopic cinema dating from the 1950's (Lipton, 1982; Spottiswoode & Spottiswoode, 1953). Much of this work was based on theoretical and mathematical analysis, with occasional references to stereoscopic depth perception by the viewers. Conversely, while there have been many perceptual studies of stereoscopic depth perception, there has been relatively little behavioural research on the relationship between the geometric parameters of a stereoscopic camera and display system and human depth perception.

In motion pictures, because objects or the camera can move, perception of stereoscopic motion in depth is important. Specifically, stereopsis (the perception of depth from the stimuli presented to the two eyes) can signal to viewers the rate of change in motion of objects, including apparent velocity and acceleration of objects in depth. Theoretically, the stereoscopic parameters of the camera configuration should affect the viewers' perception of motion in S3D content. Practically, we can ask whether, as the stereoscopic parameters are changed, there is an effect on how viewers perceive the S3D space portrayed in the image; a change in perception of space and distance between objects in the scene should ultimately affect how viewers perceive motion of objects traversing the stereoscopic space.

This chapter lays out the following important elements related to this thesis: fundamentals of the stereoscopic depth perception in humans, stereoscopic geometry, expected stereoscopic distortions based on this geometry, past psychophysical studies on stereoscopic depth perception including motion in depth, and the effects of non-linearly manipulating disparity in the S3D post-production process.

1.1 Stereoscopic Depth Perception in Humans

Binocular stereopsis or perception of three-dimensionality from binocular parallax, results from the fact that we have two laterally-displaced eyes, which capture two disparate views of the world. Visual processes in the brain unify the two images and recover information about depth.

The absolute binocular parallax of an object is the angle subtended by the baseline of the two eyes at the object. To align the eyes on an object of interest this parallax must be adjusted for. Thus, when the two eyes fixate on a point or object in space, the axis of the two eyes converge at the point of fixation; the angle at this intersection is known as convergence or vergence angle. The tangent of the vergence angle is inversely proportional to the viewing distance; when the distance between the viewer and point of fixation increases, the vergence angle decreases and vice versa. In addition, the tangent of the vergence angle is directly proportional to the interocular distance so that the larger the separation between an observer's eyes the more they must converge to fixate on a given object. The relationship between interocular distance, fixation distance, and vergence angle is captured in the equation:

$$angle = 2\arctan\left(\frac{IO}{2V}\right) \tag{1.1}$$

Retinal positional disparity refers to the distance between the images of a given object on the left and right retina. Absolute retinal disparity is defined using coordinates centered on the foveae and thus depends on fixation (convergence). Consider when the eyes are fixated on a certain point, F, at a distance V_1 with a vergence angle of α and another object, P, is located at distance V_2 with a binocular parallax of β. The retinal images of F fall on the centre of each fovea, resulting in zero retinal disparity by definition. However, unless β is equal to α, the image of object P falls on non-corresponding points on the two retinae[1]. The amount of retinal disparity is equal to the difference in binocular parallax between the object and the fixation point ($\beta - \alpha$). The horopter refers to the locus of points in space, including the fixation point, that stimulate corresponding points on the two retinae. If the object lies on the horopter, then the absolute retinal disparity is zero. Points that do not lie

---

[1] In this chapter we make the simplifying assumption that the optical centre of the eye and its centre of rotation are the same – thus the required vergence to fixate a point, its binocular parallax and its disparity when fixated at infinity are the same.

on the horopter stimulate non-corresponding locations of the two retinae, and thus have retinal disparity.

If an object is inside the horopter (such as point B in Figure 1.1), then the absolute retinal disparity of the object is referred to as crossed disparity. Crossed disparity implies that the viewers' eyes have to cross to fixate on the object or, alternatively, that the visual lines from the eyes to the object intersect or cross before the horopter. If the object is outside the horopter (such as point A in Figure 1.1), then the absolute retinal disparity is known as uncrossed disparity because the eyes must uncross to fixate the object.

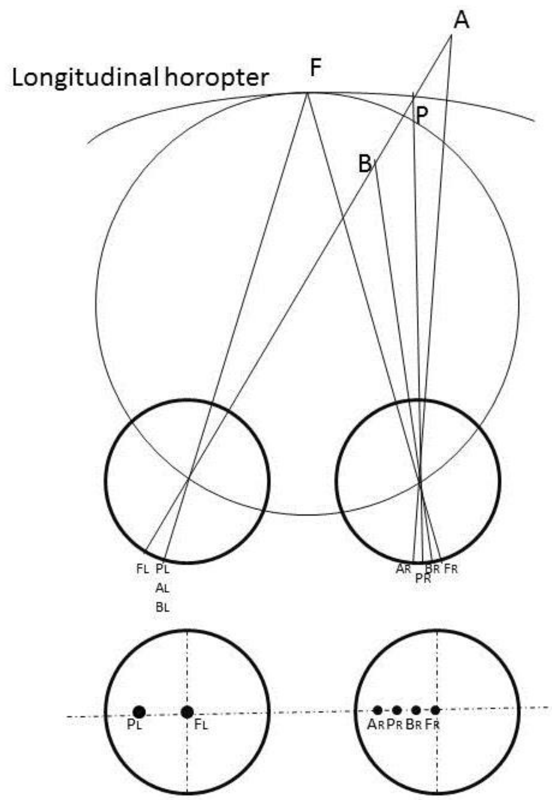

Figure 1-1

This figure illustrates relationship between different points in a scene for a person fixating at point F. The lower diagram shows the views of the retinas from the rear. Light from F falls on the centers of the two foveas, F_L and F_R. Light from point P on the horopter stimulates corresponding retinal points P_L and P_R, to the left of the center. Point A, beyond the horopter, and point B, closer than the horopter, stimulate the left eye on the same point as P. In the right eye, light from A falls on A_R, a retinal point that is more distant from the fovea than P_R and light from B falls on B_R, a retinal point nearer to the fovea than P_R. It is worth noting that point A, which lies farther than the horopter, produces uncrossed disparity. Point B, which lies closer than the horopter, produces crossed disparity. Adapted from (Hershenson, 1999)

The greater the distance between the object and the horopter, the larger the retinal disparity. Absolute retinal disparity provides stereoscopic cues to the distance between the object and the viewer, if the convergence of the eyes is known.

If there is more than one object in view, then the *relative* retinal disparity between them is equal to the difference in binocular parallax of the two objects. Note that if there is an object at the fixation point, relative disparity between the fixated object and another object will be the same as the second object's absolute disparity. The distinction between relative and absolute disparity is important because humans are much more sensitive to depth from relative disparity than absolute disparity (Howard & Rogers, 2002). If the depth between two objects is D, then the relative retinal disparity, η, varies with the inverse of viewing distance squared. If the viewing distance, V, is large compared to IO then the relative disparity can be approximated as:

$$\eta = IO * \frac{D}{V^2} \tag{1.2}$$

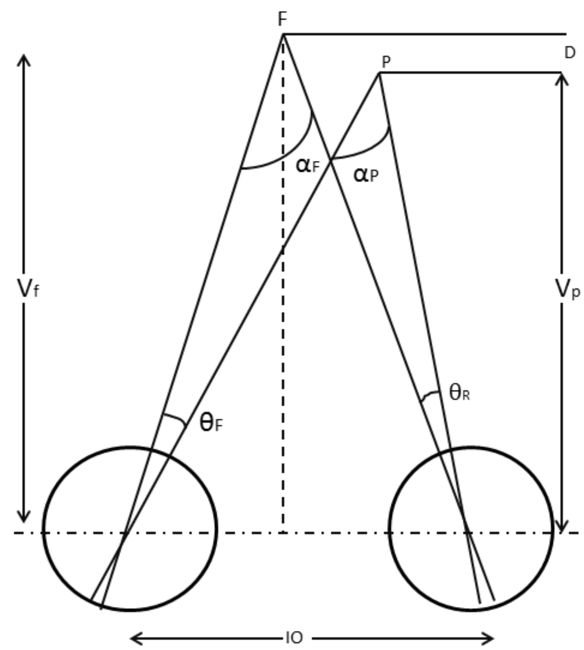

Figure 1-2

Figure 1.2: Calculation of retinal disparity, η, where IO equals interocular distance, F = Fixation Point, P = a point off the horopter, V_f = the distance to F perpendicular to the interocular axis, V_p = the distance to P perpendicular to the interocular axis, D =the difference in their distances (V_f - V_p), α_F= the apex angle of the fixation point, α_P= the apex angle of the other point, angle θ_f= the visual angle between the two points in the left eye, and θ_R= the visual angle between the two points in the right eye. Adapted from (Hershenson, 1999)**.**

1.1.1 Size constancy

The size of the retinal image of an object in space is inversely proportional to its distance. Size constancy is the ability of the viewer to derive a constant estimate of size of an object at different distances despite this variation in image size (Wallach & Zuckerman, 1963). Research has shown that binocular distance cues can be used to achieve size constancy; however, these cues including vergence fail to serve as effective distance cues for size constancy beyond a distance of about 2m (Leibowitz & Harvey, 1967). This means that at longer distances typical of most theatre situations, binocular and oculomotor cues, such as vergence and accommodation, are not sufficient for size constancy. However, in these situations, monocular distance cues in the image, distance to the screen, and binocular disparity cues in the stereoscopic image serve as possible mechanisms for perceived size and size constancy.

1.1.2 Depth constancy

As mentioned above, viewing distance is a critical variable in how the viewer perceives S3D stereoscopic content. Perceptual studies show that the viewers have the ability to account for distances when making depth estimates (Wallach & Zuckerman, 1963). This ability to take distance into consideration when perceiving depth is known as stereoscopic depth constancy. Geometry informs us that size of the retinal image of an object decreases proportionally to the object's distance from eyes; however, the retinal disparity for a given depth between two objects (or parts of given object) is inversely proportional to the square of their distance from the eyes. This change in disparity due to distance occurs for two reasons. One, the disparity results in differences of width in two retinal images and, therefore, should decrease in proportion to distance of the object from the eyes (See Figure 1.3). Also, because disparity is the product of two views of an object from different directions, the larger the difference between the two directions, the greater the disparity that result from the particular depth interval. For example, when the object is farther from the eyes, the direction between the two views is smaller, thus resulting in smaller disparity from a given depth-interval. These observations and geometric facts mean that disparity should decrease in proportion to the distance squared of the object from the eyes (See Figure 1.3 below) (Wallach & Zuckerman, 1963).

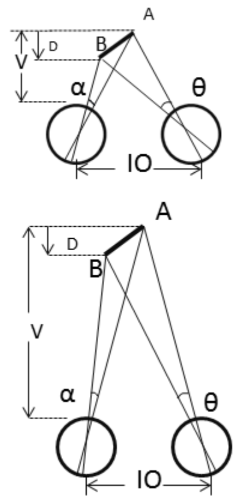

Figure 1-3

Figure 1.3: This figure illustrates the relation between stimulus depth, D, viewing distance, V, interocular distance, IO, and retinal disparity, θ- α. A comparison of two diagrams above shows the effect of V on disparity; As the viewing distance increases, the difference in direction between the two views of the eyes is smaller, thus resulting in smaller disparity (θ - α) from a given depth-interval. Adapted from (Ritter, 1977)

With perfect depth constancy, a fixed depth interval should appear to have the same depth regardless of distance from the eyes. Studies have found that depth constancy is good for depth intervals at close distances (less than 1 m) (Wallach & Zuckerman, 1963). For larger distances, especially in theatres and home viewing situations, there is partial depth constancy facilitated by rich depth monocular cues such as shading, perspective, occlusion, and binocular cues (Allison, Gillam, & Vecellio, 2009).

1.1.3 Interaction of depth and size constancy

Size constancy and depth constancy are related phenomena in that both depend on a viewers' ability to account for viewing distance. However, they are quantitatively and qualitatively different in that the size of the retinal image is inversely proportional to distance whereas the retinal disparity is inversely proportional to the distance squared. How these characteristics interact with each other will impact how the shape of the objects and their scale is perceived by the viewer.

Orthostereoscopic conditions refer to stereoscopic viewing and capture conditions that are congruent with natural human binocular vision. In an orthostereoscopic condition, the stereoscopic geometry of a real scene is reproduced on a S3D screen (Lipton, 1982) so that depth and size constancy are (ideally) maintained and stereoscopic shape perception is preserved. However, this perfect congruence to the real world space is impractical in the case of stereoscopic film. In real-world situations, when there is a change in distance, both disparity and image size change appropriately. Due to the choice of lens, IA or other parameters during stereoscopic capture, these size and disparity cues may conflict with each other and result in perceptual distortions.

1.2 Stereoscopic Parameters and Process

Because camera geometry often does not match natural binocular geometry there is a non-linear mapping between real world space and stereoscopic space presented by the S3D camera and display system. This non-linear relationship can produce distortions of space in the stereoscopic image with relative expansion or compression of space in isolated parts of the scene. Therefore, it is necessary to review the key parameters of a stereoscopic camera and display system and outline how changes in those parameters theoretically affect the projected image. Stereoscopic geometry outlines how various parameters affect the transmission of the stereoscopic image. It clearly lays out the relationship between the different parameters, so that it is easy for content producers to make changes depending on the desired effect. However, the viewers' perceptual experiences do not always match mathematical predictions. To better appreciate how perceptual experience deviates from geometric predictions, it is critical to gain a solid understanding of these geometric factors in the processes of acquisition, projection, and viewing. Each of these three phases is associated with interdependent geometric parameters that have to be adjusted to create a particular desired experience for the viewer. The most critical of these parameters are the interaxial distance, convergence, viewing distance, distance of objects, focal length, screen width, and interocular separation of the eyes. The following sections will outline how the various parameters should impact perceived depth.

1.2.1 Interaxial Distance

One of the key parameters of a stereoscopic camera and display system is the lateral distance between the centers of projection of the right and left cameras. This horizontal separation controls the disparity range of the two images and, hence, establishes the binocular parallax range of the shot. It is a critical component of how the real world depth in the scene maps to the depth in the projected image.

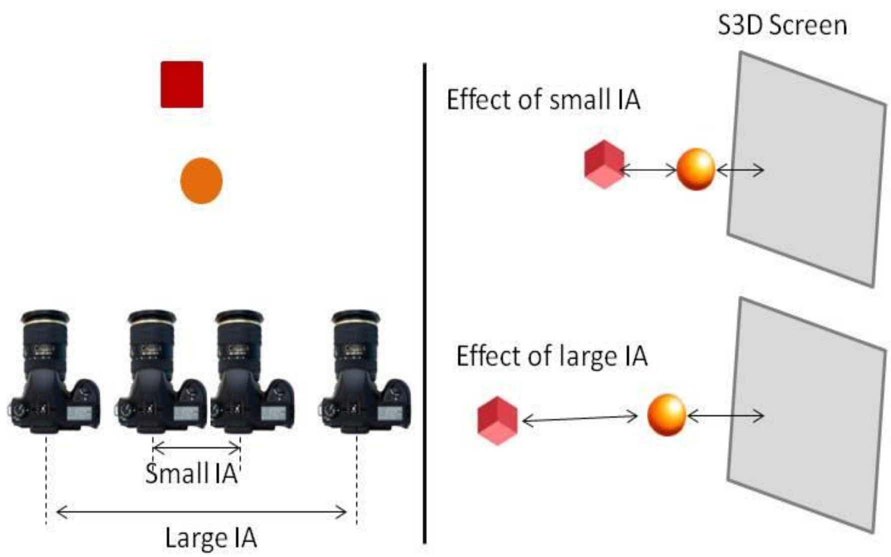

Figure 1-4

Figure 1.4a (on the left) shows a ball and a cube being captured with small camera separation (small IA) and with large camera separation (large IA). Figure 4b (on the right) shows that the predicted distance between the two objects is smaller when the small IA is used. Conversely, the apparent separation of the two objects is increased with the large IA. (Adapted from K. Benzeroual's slides, 2011)

For example, as shown in Figure 1.4a increasing the interaxial (IA) separation increases the binocular parallax between points in a scene. Thus, the perceived depth between any two points in a given scene should appear larger if filmed with a larger IA, all other things being equal. Therefore, IA determines the relationship between physical distances in the original scene, and the predicted amount of depth in the projected scene.

One of the key parameters cinematographers control on set using IA is the maximum range of binocular parallax and, therefore, the range of relative distances displayed to the viewer. In one case, Bercovitz proposed a formula for setting IA based on object distances in the scene (Bercovitz, 1998):

$$IA = d * \frac{FO*NO}{(FO-NO)} * (\frac{1}{f} - \frac{1}{a}) \qquad (1.3)$$

In the previous equation, d is the maximum desired parallax between left and right image points (i.e., the parallax range of the shot[2]). *FO* denotes the distance of the farthest object in the scene, *NO* denotes the distance of the nearest object in the scene, f is the focal length of the camera and a is the distance at which the camera is focused (by the thin lens equation the camera sensor plane should be located at $1/f - 1/a$ behind the camera nodal point).

---

[2] The parallax range is expressed in the same units as FO, f and the other distances and is the linear displacement on the camera sensor. This can easily be converted to a percentage of the image width or a number of pixels.

There has been considerable attention devoted to IA in stereoscopic cinematography because, in traditional filmmaking, once a scene is shot with a given IA the disparity range cannot be adjusted easily in post-production. Instead, changing the disparity range (hence, manipulating IA in post-production) requires depth reconstruction and view interpolation or 2D-S3D conversion techniques which are complex and costly operations (Lang et al., 2010). Currently, film production companies are developing and using software tools that allow easier manipulation of disparity in the post-production process.

1.2.2 Convergence or Zero-Parallax Setting (ZPS)

Camera convergence in a stereo camera and display system shifts the range of depth in the projected image relative to the screen plane. Therefore, the convergence point is also known as the Zero-Parallax Setting (ZPS), which is defined as the point in the scene that will be imaged on the plane of the screen. Control of ZPS is typically achieved through a toe-in configuration of the stereoscopic rig on set or through horizontal image translation (HIT). Establishing the ZPS also determines the parts of the screen that will appear to be in front or behind the screen plane. Because changing the ZPS theoretically impacts object size and depth, changes in either or both IA and convergence have implications for predicted depth from binocular disparity.

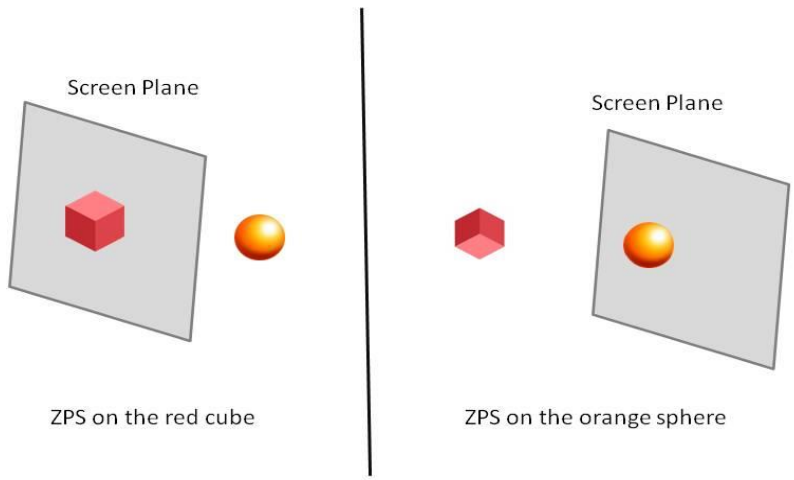

Figure 1-5

Figure 1.5: This figure illustrates how objects are viewed in relationship to each other and the screen plane based on the point at which the ZPS is set. (Adapted from K. Benzeroual's slides, 2011)

Of the two methodologies used to achieve camera convergence, shooting with a stereoscopic rig in parallel configuration and using HIT for shifting convergence is generally favored since it causes minimal distortions (Allison, 2004). In contrast, in the "toe-in" condition, the keystone distortion caused by the difference in orientation of the toed-in camera sensors produces patterns of horizontal and vertical screen disparities that can result in distorted stereoscopic depth (Allison, 2004; Woods, Docherty, & Koch, 1993).

Nevertheless, the “toe-in” method has some advantages that render it useful for content creators. For example, with the cameras’ limited field of view and non-zero IA, there are regions of space that are not captured by one or both cameras and toeing in the cameras can allow these areas to be imaged. The “toe-in” camera configuration also offers the possibility of centering the target on the camera optics (thus reducing camera distortions).

1.2.3 Focal Length

The focal length of a lens refers to the distance from the optical center of the lens to the principal focus. The principal focus of a converging lens is the point to which the rays parallel to the principal axis converge after refraction by the lens and is where the camera sensor should be placed when focused at infinity. When focusing at nearer objects the lens should be located further away than the focal length as described by the Bercovitz equation (Equation 1.3). The angle of view refers to the amount of scene that is captured on the image sensor, with a wide-angle lens capturing larger amount of the scene. A “normal lens” refers to the condition when the image captured by sensor matches the field of view of the displayed image so that the lens reproduces natural human field of view and makes the scene look natural in terms of depth of field, movement, and perspective of the scene. Focal lengths shorter than the ‘normal’ lens result in a wide angle of view. Conversely, a large focal length reduces the angle of view, diminishing the area of scene being captured; this is known as a long focus or telephoto lens (Lipton, 1982).

As noted above, the IA has implications for how the size and depth of the object is perceived. These predicted effects are enhanced or reduced depending on the particular lenses used. Long focal length lenses are often used to film distant scenes by narrowing the field of view and magnifying the far objects. It is important to note that when shooting with a telephoto lens, the distance between the object in view and camera is typically perceived to be closer and the scene appears flatter than the actual geometry of the physical configuration. This happens because long focal lenses magnify the far object and compress the distances or perspective between objects. The human visual system compensates for this distortion by maintaining a consistent shape and form and interpreting the scene to be closer to the observer than implied by geometry (Benzeroual, Wilcox, Kazimi, & Allison, 2011). The perspective effects of these distortions can impact the interpretation of the disparity and geometry of stereoscopic images.

1.3 Aspects of projection that impact stereopsis

1.3.1 Screen Parallax

One of the critical parameters of the projection geometry is the screen parallax. Screen parallax is similar to retinal disparity but is defined in terms of the projected images on the screen not the retinae. Crossed or negative parallax occurs when the screen position of the image of a point P in the left-eye view lies to the right of its position in the right-eye view. Crossed screen parallax (see Figure 1.6) specifies a stereoscopic image located in front of the display screen. Conversely, uncrossed or positive parallax occurs when the left-eye image of P is to the left of its right-eye image on the screen and specifies a point lying behind the screen. As discussed above, zero parallax setting (ZPS) refers to the point in the stereoscopic scene that will be imaged on the plane of the screen (and therefore have zero screen parallax). The ZPS can be modified during image capture or in post-production to adjust which part of the image will appear in front or behind the screen display.

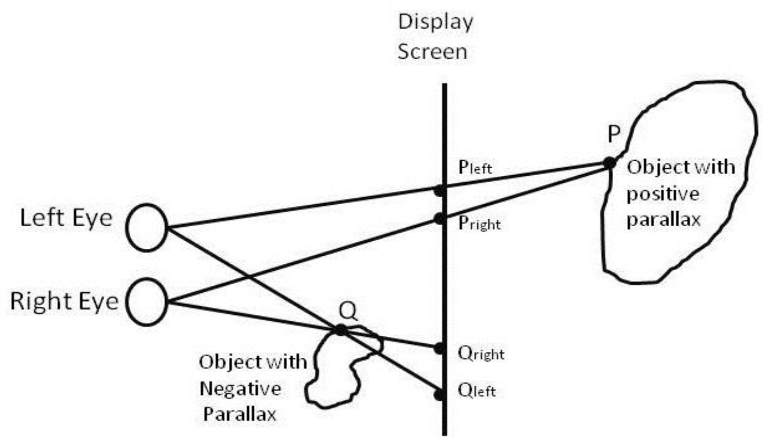

Figure 1-6

Figure 1.6: This figure shows the relationship of image points on the screen to the projection of objects behind or in front of the screen. An object with a positive parallax is seen behind the screen with image point, P, resulting from uncrossed screen parallax (where P_{left} is to the left of P_{right}). Conversely, an object with negative parallax is seen in front of the screen with image point, Q, resulting from crossed screen disparity (where Q_{right} is to the left of Q_{left}).

To review the projection geometry, let's consider V to be the perpendicular viewing distance from the viewer's eyes to the screen plane, which is also the ZPS plane in that the points on the screen have zero parallax. IO is the distance between the two eyes; S_p is the separation between left and right homologous points on the screen, and D is the depth or distance between the screen plane and the stereoscopic image (D is negative for crossed parallax). By similar triangles, $Sp = D * IO / (V + D)$.

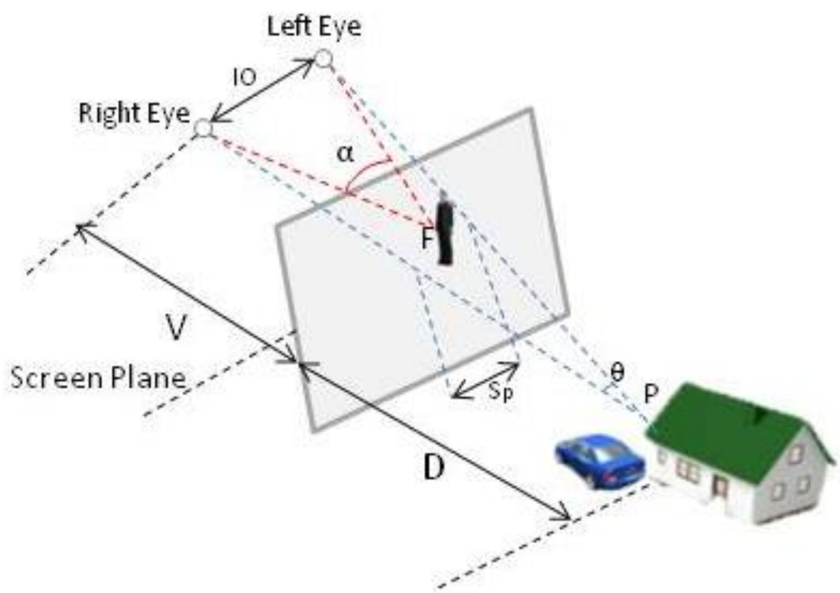

Figure 1-7

Figure 1.7: This figure illustrates the geometry of stereopsis with V: viewing distance; D: relative distance between an object P and the fixation point F; IO: interocular distance; S_p: is the on-screen

parallax of the point P. Using similar triangles Sp = D * (IO)/(V + D), where D is negative for crossed screen parallax. Adapted from (Benzeroual, Allison, & Wilcox, 2012)

It is important to note that the screen parallax is not the same as retinal disparity. The amount of retinal disparity for a given amount of screen parallax depends on the viewing parameters (interocular distance, viewing distance, convergence, etc).

It is a useful exercise to review how the different display parameters, such as screen parallax and viewing distance affect the predicted depth experience for the viewer. When there is positive screen parallax then as the S_p is increased, depth relative to the screen, expressed as $D = V \cdot Sp/(IO - Sp)$, is increased. When S_p equals IO then the predicted depth is infinite. The converse is true when the S_p is decreased. For negative screen parallax, as negative S_p is increased, the depth in front of the screen increases. For this case, as S_p approaches infinity the predicted D approaches the viewing distance placing the object infinitesimally close to the viewer. Also, it is important to note that when the magnitude of negative screen parallax, S_p, is equal to the interocular distance of the viewer, the depth of the imaged point is predicted to be half the viewing distance. Thus, the same magnitude of screen parallax (S_p) with opposite sign indicates infinite depth beyond the screen but a depth of only half the viewing distance in the negative parallax case. Thus, the mathematical models show that there is an asymmetric relation between predicted depth from negative and positive screen parallax in that the ratio of depth to viewing distance can never exceed 1.0 with negative parallax, yet this ratio can be infinite in the case of positive parallax.

Viewing distance also impacts how the viewer perceives depth in the stereoscopic image. Since the angular size of the linear separation between the left and right images decreases with viewing distance, the depth from a given screen parallax varies proportionally to viewing distance (rather than viewing distance squared as in the angular disparity case). If the viewing distance is halved, the depth for a given parallax will decrease by two; conversely, if the viewing distance doubles, then the depth will double

1.3.2 Screen Size

Diversity of screen media has made adaptability of S3D content to varying screen sizes one of the most pressing concerns amongst film producers and distributors. While 2D media can be easily adapted through straightforward interpolation methods, in S3D content the underlying stereoscopic parameters may need to be readjusted for different screen sizes to control depth and avoid distortion.

Currently, S3D movies are displayed on a variety of screen sizes which include IMAX 3D (~18m x 24m), traditional cinema 3D (~12m x 5m), TV 3D (54" diagonal), computer screens (22" diagonal) and mobile devices (5" diagonal). The wider the range of screen sizes targeted for S3D content the more constrained the production and post-production processes become. Once an S3D image has been acquired, magnification of the image (depending on the screen sizes) impacts the screen parallax. Both geometric studies by Spottiswoode and experiential work by Lipton (Lipton, 1982; Spottiswoode & Spottiswoode, 1953) showed that, all other aspects being equal and because of the role of the IA in controlling the parallax range of the shots, the IA chosen restricts the range of screen sizes on which the content can be viewed optimally and comfortably.

Typically a larger screen is viewed from a larger distance; standardized viewing recommendations maintain a constant viewing angle by setting viewing distance proportional to screen size (according to THX (Thx.com) and SMPTE (smpte.org) viewing recommendations). When the viewing distance is increased and viewing angle is constant, S3D effects are amplified. In this case, amount of predicted depth is increased (viewing distance effects are considered in the next section).

On the other hand, if the positive parallax is too high, the viewers' eyes need to diverge to fuse the images of far objects. For example, let us assume that the image of an object on small screen has a parallax that equals to the viewer’s IO. When that same image is projected on a large screen the parallax is increased (and, therefore, becomes greater than the viewer’s IO) due to magnification. In this case, the viewer’s eyes will be forced to diverge to fixate and fuse the object. Because of a TV screen’s smaller size than a large cinema screen, a given image presents a smaller parallax. Therefore, there is a higher limit on positive parallax before divergence issues are encountered; this means that S3D content optimized for TV may have excessive disparity for the cinema.

1.3.3 Viewing Distance

The viewer's distance from the screen, significantly impacts how the viewer perceives the stereoscopic image. In natural binocular vision, the relationship between disparity and viewing distance is quadratic: $\eta \approx \frac{IO*D}{V^2}$which implies that as the viewing distance (V) increases the retinal disparity for a given depth interval is reduced quadratically (Howard & Rogers, 2002). However, this relationship does not describe the predicted depth when the images are viewed on a stereoscopic display. The reason is that screen size and screen distance are often inter-dependent such that smaller screens are often viewed from a closer distance than large ones. If the angular size of the screen subtended at the eye is held constant then, since the screen parallax in pixels does not change, the binocular disparity (the angular subtense of the distance between left and right image points on the screen) is also kept constant while the viewing distance is changed. Thus, the amount of depth for a given S3D image is smaller on a small, near display than on a large, distant display.

1.4 Perceptual Distortions

As stereoscopic content is captured and displayed, real world space is translated / mapped to portrayed space. As discussed above, this mapping is not a linear one. After the display phase of stereoscopic transmission, the S3D content enters the perceptual space of the viewer, which means that distortions introduced through capture and display have implications for how the viewer perceives these distortions. It is, therefore, important to review exactly what parameters introduce which kinds of geometric distortions and how those distortions are perceived by the viewer.

1.4.1 Puppet Theatre Effect

The reduced size or miniaturization of objects in the scene as a result of excessively large IA (hyperstereopsis) is also known as the puppet-theatre effect. In the puppet theatre phenomenon, the foreground objects appear to be small, but properly scaled, meaning they do not show shape distortions. Miniaturization is direct by-product of the IA used to capture content; as mentioned above, as the IA is increased, so does the binocular parallax between points in the scene thus increasing disparity within objects and between objects and the background. Due to the large IA, if the objects were perceived at their normal scale then the mismatch between disparity and size would lead to exaggerated depth and distorted shape (Yamanoue, Okui, & Okano, 2006). The visual system appears to avoid this distortion by interpreting the objects in the scene as nearer and smaller scaled versions. As depth from disparity decreases with decreasing distance faster than size does, this allows the visual system to assume an apparent distance that produces the perception of proportionate and scaled objects (Benzeroual, Allison, & Wilcox, 2011)

1.4.2 Cardboard effect

Another perceptual artefact as a result of perceived size and perceived disparity mismatch is the cardboard effect. In this artefact viewers perceive objects in the image to be flattened like cardboard cut outs. It typically appears as if the stereoscopic image is divided into discrete depth planes and curiously the depth between the layers often appears less compressed than within the objects.

Past research has related this artefact to the fact that whereas the size of the image of an object is proportional to the inverse of its distance ($1/V$), the disparity within the object is inversely proportional to the square of its distance ($1/V^2$)(Howard & Rogers, 2002). When objects in a stereoscopic image appear to be too near then their apparent depth is compressed relative to their apparent linear size, thus making the objects appear to be flattened (Ijsselsteijn, Seuntiens, & Meesters, 2005). This phenomenon is related to the puppet theatre effect with the exception that there is no normalization that allows the objects in the image to appear proportionate (Benzeroual, Allison, et al., 2011).

An object can appear to be seen as ‘too near’ when S3D content is shot from shooting distances which are much greater than the viewing distance. When viewing the S3D content, the vergence, accommodation, and other cues such as perspective denote a shorter viewing distance. If the image is interpreted in accordance with this shorter distance, then the cardboard cutout distortion can arise from volumetric depth being scaled down with viewing distance. There have been research attempts to eliminate this effect by synchronizing the viewing conditions with the shooting conditions, resulting in elimination of the distortion in some cases. This suggests that one of the other reasons for this distortion may be cue conflicts arising from differences between shooting and viewing conditions (Yamanoue et al., 2006).

The use of long focal lens can also lead to the cardboard effect. The use of telephoto lens in magnifies background objects relative to foreground images and, consequently, compresses apparent distance between objects in the scene and between the camera and the object (Yamanoue et al., 2006). This magnification reduces both perspective and disparity (differential perspective) between left and right images (Benzeroual, Allison, et al., 2011). As stated above, the viewing distance between observer and screen plane in S3D content is generally smaller than the capture distance between camera and shooting target. The combination of all these factors results in acute foreshortening of objects in the scene and compression of depth in the scene across different layers (Ijsselsteijn et al., 2005).

1.4.3 Roundness

The roundness factor is the ratio between perceived width of an object and its perceived depth so that it measured by depth / width. In orthostereoscopic situations in which viewing parameters are synchronized with shooting parameters, perceived depth should be similar to real depth of an object and the roundness factor should equal 1 (Mendiburu, 2012). When the roundness factor equals to 1, it means that a sphere is seen perfectly as a sphere. When it is smaller than 1, it appears to be flattened with small depth. On the other hand, if it is larger than 1, it appears to be elongated in depth like a rugby ball.

In simplified terms, four factors have to be balanced to have perfect stereoscopic roundness. In terms of the viewing conditions, those factors are the interocular distance, IO, and viewing distance, V. With regards to photographic situation, it is the convergence distance, C, and the interaxial, IA (Wattie, 2012).

$$V = C * IO/I \qquad (1.4)$$

From these factors, it is obvious that the only way roundness can be preserved is by photographic configuration matching display configuration.

It is important to remember that although filmmakers may know the eventual display parameters, it is difficult to shoot strictly according to the desired parameters as shooting conditions vary depending on the type of project. For example, shooting live events is different from shooting scripted material (which is sometimes more conducive to being constrained than live productions) (Mendiburu, 2012).

From the equations above, it is possible to conclude that as the viewer moves closer to the screen (i.e., by choosing a nearer seat in a theatre), the objects in the scene will appear to be more compressed (with roundness less than 1); conversely, moving farther away from the screen results in perceptual stretching of the object (with roundness factor greater than 1) (Benzeroual, Allison, et al., 2011).

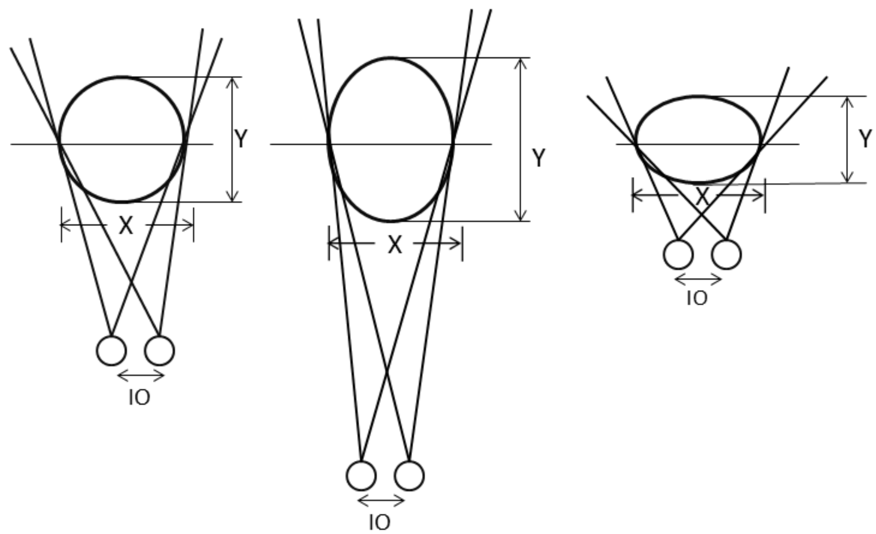

Figure 1-8

Figure 1.8: This figure illustrates roundness as it relates to the viewing distance for a fixed size image/screen. The predicted size of depth, Y, increases as viewing distance increases. The converse is true when the viewing distance decreases where size of depth, Y, decreases. Adapted from (Benzeroual, Allison, et al., 2011)

1.5 Non-linear Depth Scaling in S3D

In addition to the viewing experience related to screen sizes, filmmakers often manipulate depth in the post-production processes to optimize viewer comfort and to enhance the use of S3D storytelling as an artistic and narrative device. Because of the complex interactions among the various parameters, the S3D film industry is constantly devising new techniques and best practice rules concerning the production and display of S3D movies.

Developing and implementing these rules requires understanding of how to control disparity during filming and post-production. Currently, an active area of research entails developing algorithms to manipulate the range of disparities, and hence the predicted apparent depth of a scene, in postproduction so that content can be adapted for various display conditions (Lang et al., 2010; Trivedi & Lloyd, 1985; Wang & Sawchuk, 2008). Flexible control over scene depth in postproduction is also an important artistic tool to create a desired emotional experience for the viewer.

Setting the disparity range of the portrayed scene, for instance by varying IA, usually entails a trade-off between controlling the maximum disparities in the scene while maintaining depth and volume in the subject. When the disparity is scaled or introduced globally through IA manipulations or in postproduction across the scene, it can result in undesired artefacts and distortions such as flattening of foreground objects. In order to provide a flexible solution to disparity mapping, non-linear depth scaling algorithms can be used to manipulate depth non-linearly across the screen so some objects can be made to show more volume than others (Lang et al., 2010).

Depth manipulation includes operations that introduce nonlinear changes across disparity ranges. Suppose that we have a scene with people or animals in the foreground with the background of the forest. Also suppose we want to make the changes to the depth such that the foreground becomes more voluminous while the background stays the same depth. In general, the changes in disparity ranges have to do with changing the disparity interval, aka depth budget, from a given range to the desired range (Lang et al., 2010). One of the critical challenges of native stereoscopic production is that once content has been captured, the only variable that can be easily altered in the content is the effective convergence through HIT. In the example above, if the disparity ranges are modified uniformly, then the depth changes will compress or expand both foreground and background objects. There are tools in the post-production process that can nonlinearly warp depth such that it is possible to give foreground objects more volume while compressing the background. These are node-based programs such as BlackMagic Fusion which uses nodes to represent effects, filters, and other processes (including effects using depth maps) to build more complex visual effects. Originally, the left image is constructed with application such as Maya 3D, which is imported in Fusion to create the right image using depth maps. The depth maps are then used to adjust the function between depth in (distance from the camera in the original stereo pairs) and depth out (the geometrically predicted distance from the viewer of the display). This leads to creation of a new stereoscopic image based on the depth maps and the new depth out function.

1.6 Motion in Depth

1.6.1 Target Vergence, Relative Disparity, and Looming

Research studies show that there are multiple sources of visual information that aid in perception of motion in depth. Some of the key sources of information include target vergence (which the absolute binocular parallax of the target with the simplifying assumption of coincidence of the optical and rotational centres of the eye), relative disparity, and retinal image size (Brenner, Van Den Berg, & Van Damme, 1996).

As an object approaches us, its retinal image gets bigger. The changing image size of an object as it approaches or recedes relative to its observer is referred to as looming. Although image size provides no information about the distance of a stationary object, changes in image size provide a robust perception of motion in depth. The changing image size gives an indication of the change in distance over time.

Changes in absolute disparity provide the main cue for changing vergence. When the eyes are stationary, the disparity between the left and right images of the object changes as the object moves in depth. In contrast, when the convergence of the eyes changes to fuse and track the images of the moving object, the absolute retinal disparity of the target is zero. For an interocular distance, IO, and a target at distance V, the target vergence, θ, is defined by:

$$\theta = 2\arctan\left(\frac{IO}{2V}\right) \qquad (1.5)$$ (Howard, Fujii, & Allison, 2014).

The relative disparity between two objects at varying distances is the difference in their binocular parallax. When an object moves in depth relative to a stationary reference the relative disparity change is a reliable cue to motion in depth.

Looming, change in target vergence, and change in relative disparity can all influence the perceived velocity of motion in depth and the perceived position of an object in space. However, it has been shown that sources of information for perceived distance are not always the same as those for perceived motion. For example, Brenner et al (Brenner et al., 1996) reported that changing retinal image size impacts perceived motion in depth more than perceived change in distance. In their experiment, the target's perceived velocity depended on its changing image size, which was indicative of the ball's motion in depth. They found that when target vergence and relative disparity were eliminated, the perceived distance indicated by looming increased, but it didn't result in faster perceived motion.

Perceptual studies by Brenner, van den Berg, and van Damme (1996) explored the effects of looming, vergence, and relative disparity on perception of motion in depth. They found that holding one cue constant, like looming, reduced the perceived velocity of an approaching object while other cues were changed. When only one cue was changed and others were held constant, looming produced highest perceived velocity, changing only relative disparity produced less perceived velocity, and changing only target vergence produced no perception of motion in depth (Brenner et al., 1996). The latter corresponds to the findings of Erekelens and Collewijn and Regan et al. who found no changes in perceived distance were produced by vergence eye movements made to track the changing absolute parallax of large textured surfaces (Erkelens & Collewijn, 1985; Regan, Erkelens, & Collewijn, 1986). Gonzalez et al. (González, Allison, Ono, & Vinnikov, 2010) argued that this lack of perception of motion in depth from changing target vergence was probably due to cue conflict with constant looming information since changing target vergence produced motion in depth when the conflicting looming information was weak or absent (see also Howard et al., 2014). They found that looming in random-dot displays produced stronger motion in depth percepts than disparity modulation and, in presence of looming, perceived motion in depth is always in the direction of the looming cue. Motion in depth from looming produces depth that is equivalent to perceived motion in depth by combined effects of looming and disparity changes (González et al., 2010).

Perceptual studies have often made use of compound stimuli, which includes a moving stimulus situated with a stationary reference, to test for effects of relative disparity. Gonzalez et al. used both single and compound stimuli consisting of a red dot and a 50% green and black random-dot texture together and individually to study cue conflict between disparity change and looming in the perception of motion in depth. They found that when motion in depth was perceived by disparity in one or both elements of a compound stimulus, the relative disparity was much more effective than absolute disparity. With looming held constant, relative disparity resulted in stronger motion in depth changes (González et al., 2010).

Previous experiment's findings were corroborated by a subsequent experiment by Howard, Fujii, and Allison on interaction between cues to visual motion in depth (Howard et al., 2014). Howard et al.'s experiment also used stimuli that consisted of a small central dot and a textured surface moving to and fro in depth along the midline. The stimuli comprised both a moving stimulus and fixed reference. The experiment tested for various conditions such as the role of changing looming, target vergence, and relative disparity in the perception of motion in depth. For the looming only condition, strong motion in depth was perceived for textured stimuli and weak motion in depth was seen with the dot stimuli. For the vergence only condition, strong motion in depth was sensed for the dot-alone stimulus (without the reference) but not for the textured-alone stimulus (without the reference). Adding the fixed reference stimulus increased perception of depth from target vergence and more depth was perceived in the dot-relative and texture-relative than in the dot-alone and texture-alone conditions.

In the condition where looming signals were in the same direction as vergence signals, motion in depth was seen in the appropriate direction. However, in cases where looming signals were opposite to the vergence signal, the perceived depth depended on the stimulus. They found that the cues dissociated which was further confirmed by the finding that in the presence of the fixed reference, people could perceive motion in depth in the direction of looming or in the direction of relative disparity.

The above-mentioned experiments suggest that motion in depth is a product of various sources of information including target vergence, relative disparity, and looming. These sources of information interact with one another to various degrees to create the perception of an object moving in space through depth.

1.6.2 Ground Plane Information

Another factor that influences how people see objects in depth is ground plane information. The significance of ground surfaces in the perception of relative distances between objects in 3D was suggested by Al-Hazen about a 1000 years ago (Al-Hazen, 1989), who proposed that there has to be an array of "ordered, continuous bodies" between objects and eyes to perceive distance effectively (Bian, Braunstein, & Andersen, 2006). Gibson (1946) emphasized the role of a continuous ground surface in 3Dperception of distances by stating "The problem of three dimensional vision, or distance perception, is basically a problem of the perception of a continuous surface which is seen to extend away from the observer" (Gibson, 1950). He further suggested "that there is literally no such thing as a perception of space without the perception of a continuous background surface" (Gibson, 1950).

Subsequent studies have also confirmed that the ground plane plays an important role in distance judgements between objects. Bian, Braunstein, Anderson (Bian, Braunstein, &

Andersen, 2005) explored the role of different surfaces, including a ground surface, a ceiling surface, and the sidewalls, in the perception of relative distances of objects in a 3D scene. Participants viewed a scene with two vertical posts either between the ground and ceiling surface or with horizontal posts between the sidewalls and discriminated which of the posts was nearer (Bian et al., 2005). When the two surfaces were a ground surface and a ceiling surface, the observers showed preference for making judgments according to the optical contact information of the ground surface; they called this phenomenon the ground dominance effect. They further discovered that the ground dominance effect did not depend on the height of the posts in the scene and that the dominance effect decreased as the scene was tilted away from the ground-ceiling orientation (Bian et al., 2005).

Furthermore, Ni, Braunstein, & Andersen investigated the role of optical contact (the contact between the image of an object and a background surface in a 2-D projection) and other cues, such as motion parallax, shadow, and occlusion on perceived distances of objects in 3D scenes (Ni, Braunstein, & Andersen, 2004, 2007). They found that optical contact took precedence over motion parallax when only one object was in the scene whereas when there were multiple moving objects in the scene, the optical contact information was outweighed by motion parallax. They also found that both optical contact information and information provided by shadow, especially in moving scenes, helped determine the perceived distance of an object (Ni et al., 2004, 2007). The displacement between an object and its cast shadow in an image offers crucial source of visual information about spatial layout of objects. Kersten, Mamassian, & Knill (1994), found that cast shadow information dominated many other sources of information for inferring 3D object motion, including the viewer's assumption of constant object size and a general viewpoint. They also found that human visual system assumes a stationary light source in the perceptual processing of shadow motion (Kersten, Mamassian, & Knill, 1994).

1.6.3 Perception of Acceleration in Motion in Depth

As mentioned above, real world objects can also change in speed / velocity, that is they can accelerate or decelerate. While people often misperceive smooth accelerated motion as constant velocity they are able to detect high rates of changes in speed and velocity (Gottsdanker, Frick, & Lockard, 1961). Gottsdanker et al. asked participants to distinguish between accelerating and constant-velocity motion of a target. They found that percentage of correct response increased with increasing mean velocity and with decreasing presentation time. They found that acceleration appeared to be detected by comparing velocities at the beginning and end of the movement (Gottsdanker et al., 1961). More recent studies have investigated whether visual judgments of acceleration could be used for the interception of moving targets. Brouwer, Brenner, & Smeets used differential judgment tasks of two successive dots (one accelerating and one decelerating) and absolute judgment task to determine how well subjects can detect acceleration when the presentation time is short (Brouwer, Brenner, & Smeets, 2002). In agreement with earlier studies they found that subjects did not detect the acceleration itself but detected the change in velocity between the beginning and end of the presentation. They found that velocity change of

25% was needed to detect acceleration (Brouwer et al., 2002). The most recent studies in the field have investigated the effects of direction on the ability to detect acceleration in more complex motions, such as radial optic flow while controlling dot size, speed, and density and manipulating radial direction (Mueller & Timney, 2014). They found that observers were better at detecting acceleration when viewing radial contraction rather than radial expansion.

Research has shown that there are theoretically three ways in which the visual system can code for changes in disparity, which then impact the perception of motion in depth (for review see Allison & Howard, 2011). First, there is the change-of-disparity (CD signal) which registers the change in binocular disparity over time and, consequently, affects perception of motion in depth. Second, there is the opposite motion of the two images (left and right) which is registered over time to produce interocular velocity difference (IOVD) signal that aids in perception of motion in depth. The third possibility of how perception of motion is registered is through specialized detectors which are sensitive to changing disparity in binocularly matched features. Significant amount of work has been done to isolate and dissociate these binocular signals (Allison & Howard, 2000; Cumming & Parker, 1994; Czuba, Guillet, & Cormack, 2013). The CD signal entails that the disparity is registered first at each instant and the temporal derivative (d/dt) of disparity codes for motion in depth. For the IOVD signal, the motion of each image is first registered followed by interocular difference in motion, which then codes for motion in depth (Allison et al., 2011). In the last case, there are specialized disparity sensors that detect changing disparity in absence of instantaneous disparity signals.

1.7 Summary

As mentioned in the introduction above, manipulation of depth-to-parallax mapping occurs by adjustment of the interaxial distance or through stereoscopic warping, resulting in theoretical non-linear mapping between real and stereoscopic space. These manipulations have become key standards in the stereoscopic film industry to produce and use depth as an important creative element. Many tools and apps are now available on the market to help stereoscopic creators achieve the optimal interaxial distance or stereoscopic depth remapping function to create the desirable depth experience. For this reason, it becomes relevant and necessary to analyze whether these manipulations can result in undesirable or unpredictable artifacts for the viewers. Furthermore, movement and motion are inherent part of the film process both in capture and post-production. Any factors that impact how space and distances between points in the stereoscopic scene are viewed will theoretically have an impact on how objects are perceived to traverse that stereoscopic space. It is in this spirit that this thesis investigates how viewers' perception of motion, specifically acceleration, in stereoscopic 3D system is affected by stereoscopic filming parameters such as interaxial distance and stereoscopic warping. Consequently, this study is an amalgamation of topics related to human binocular vision, stereoscopic geometry, motion in depth, and non-linear disparity mapping. The studies in the following chapters will investigate how viewers perceive motion in depth in stereoscopic content and whether change in interaxial distance or post-production techniques such non-linear disparity mapping affect this perception. It is also worth noting that although this thesis studies manipulations consistent with conventional stereoscopy and stereoscopic content creation guidelines, more extreme variation in parameters might be used in creative contexts to deliberately create distortions.

2 Motion in Depth Constancy in Stereoscopic Displays

2.1 Chapter Summary

In a stereoscopic 3D scene, non-linear mapping between real space and disparity could produce distortions when camera geometry differs from natural stereoscopic geometry. When the viewing distance and zero screen parallax setting are held constant and interaxial separation is varied, there is an asymmetric distortion in the mapping of stereoscopic to real space. If an object traverses this space at constant velocity, one might anticipate distortion of the perceived velocity. To determine if the predicted distortions are in fact perceived, we assessed perceived acceleration and deceleration using an animation of a ball moving in depth through a simulated environment, viewed stereoscopically. The method of limits was used to measure transition points between perceived acceleration and deceleration as a function of interaxial and context (textured vs. non-textured background). Based on binocular geometry, we predicted that the transition points would shift toward deceleration for small and towards acceleration for large interaxial separations. However, the average transition values were not influenced by interaxial separation. These data suggest that observers are able to discount distortions of stereoscopic space in interpreting the object motion. These results have important implications for the rendering or capture of effective stereoscopic 3D content.

2.2 Introduction

The recent commercial success and interest in stereoscopic cinema has spurred renewed research into the tools, processes, and perceptual experiences of stereoscopic film. In almost all cases the stereoscopic imagery differs from what the viewer would experience when viewing the same scene directly. While the specific geometry of stereoscopic 3D (S3D) film is clearly documented (Lipton, 1982; Spottiswoode & Spottiswoode, 1953; Woods et al., 1993), there have been fewer studies of how display and camera parameters affect the viewer's perceptual experience. Since, for a variety of reasons, camera geometry does not match natural binocular viewing geometry there is a non-linear mapping between real world space and stereoscopic space captured by the S3D camera system. This non-linear relationship can produce distortions of space in the stereoscopic image with relative expansion or compression of space in isolated parts of the scene (Woods et al., 1993). Although these geometric distortions of space are tolerable to viewers to a certain degree, they can also contribute to an "unnatural" look or feel of space in the image (Ijsselsteijn, De Ridder, & Vliegen, 2000).

It is also possible that distortions of space that are imperceptible in static scenes may become more visible (even disruptive) if a moving object passes through this region. The purpose of this paper is to evaluate how an object moving through stereoscopic space (with varying monocular cues) is perceived by the viewer.

Key parameters in stereoscopic cinematography include interaxial distance (IA) and zero-parallax setting (ZPS), along with other variables such as viewing distance, sensor size, screen size, and lens choice. ZPS refers to the point in the scene that will be imaged on the plane of the screen (and therefore have zero screen parallax). The ZPS can be manipulated by toe-in of the cameras during image capture, and/or via horizontal image translation during the post-production process. Adjusting the ZPS will also determine which parts of the scene appear in front of or behind the screen plane, and as outlined below, this has implications for the amount of depth predicted from binocular disparity.

Interaxial, or the lateral separation distance between the cameras, establishes the binocular parallax range of a shot. Like the separation of the eyes in the human visual system, increasing the interaxial separation increases the binocular parallax between points in a scene, even though their physical separation remains constant (and vice versa). Therefore, IA determines the relationship between distances in the scene, and the predicted amount depth in the image. In stereoscopic cinematography considerable attention has been devoted to IA primarily because once a scene is shot with a given IA, it cannot be adjusted easily in post-production (unlike the ZPS, which simply requires adjustment of the relative horizontal position of the images). Instead, changing the amount of parallax requires depth reconstruction and view interpolation or 2D-3D conversion techniques (Lang et al., 2010) which are complex and costly operations.

2.2.1 Depth Distortions in Stereoscopic Space

Spatial distortions should be minimal in orthostereoscopic conditions. However, shooting under orthostereoscopic conditions is not always practical or desirable and in any case, can only hold for a single position in the room or theatre. Because perceived depth is a function of several different stereoscopic parameters, spatial distortions can be introduced in any of the stages of stereoscopic transmission including capture, display, and /or viewing. The mapping of scene depth to screen parallax between two objects is a nonlinear function of their distance scaled by IA; a given depth interval projects to a larger relative parallax at near compared to far distances in the scene.

This non-linearity also applies to the natural stereoscopic system of the viewer, where the interocular separation between the eyes (IO) plays a role analogous to the camera IA. For a viewer with their head aligned (the baseline between the eyes parallel to the screen and centered on the image) the relationship between screen parallax, S_p, and predicted (re-projected) depth relative to the screen, d, for objects lying along a line perpendicular to the baseline through the centre of the screen is:

$$S_p = \frac{IO \cdot d}{V + d} \qquad (2.1)$$

where V is the viewing distance. The relative parallax between two objects is the difference in their screen parallax:

$$S_{p_{2\text{-}1}} = \frac{IO \cdot \left(d_a + \frac{d_{2-1}}{2}\right)}{V + d_a + \frac{d_{2-1}}{2}} - \frac{IO \cdot \left(d_a - \frac{d_{2-1}}{2}\right)}{V + d_a - \frac{d_{2-1}}{2}} \simeq \frac{IO \cdot d_{2-1}}{V + d_a} \quad (2.2)$$

assuming the relative depth $d_{2\text{-}1}$ between the objects is small relative to their average distance from the viewer (V+d_a). Although this relationship is non-linear it is the natural geometry of stereopsis and, when possible, the viewer should account for distance to perceive depth without distortion. As discussed in the introduction, this ability is known as depth constancy and has been shown to hold for stereopsis, at least in part, at both near and far distances (Allison et al., 2009; Wallach & Zuckerman, 1963). However, when the scene has been captured and displayed with stereoscopic parameters that do not match the viewer's, the viewer must interpret the nonlinear transformation between stereoscopic space and real space acquired by the cameras based on his / her natural interocular separation (IO). The differences in the nonlinear relationship caused by differences between IA and IO should be perceived as depth distortion if interpreted geometrically. Figure 2.1 below shows how relative depth between pairs of sample points in the real world maps nonlinearly to stereoscopic space with varying IA but fixed ZPS and screen distance.

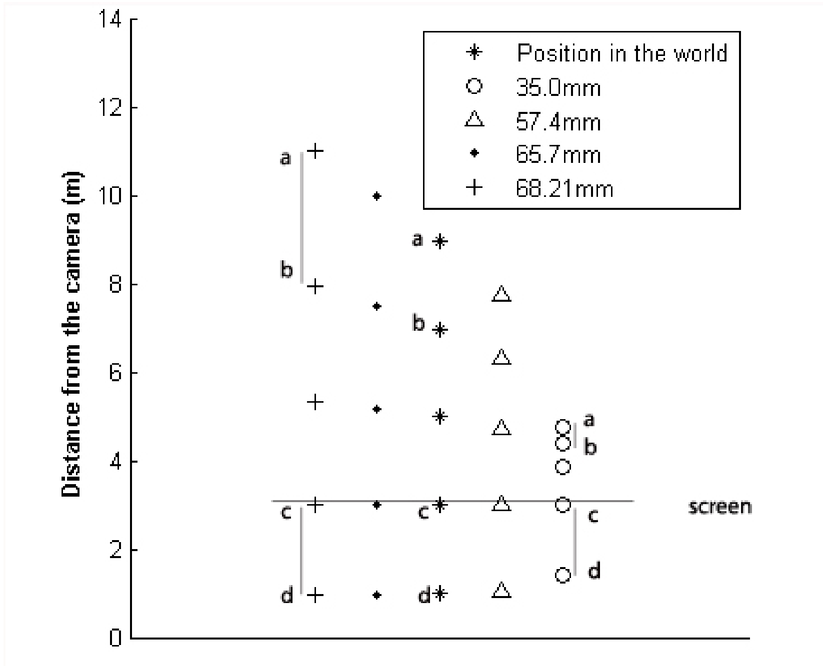

Figure 2-1

Figure 2.1 shows how points in the real world map to positions in predicted stereoscopic space according to IA assuming a matched camera and screen field of view and camera convergence at 3 m (the screen distance). Stereoscopic images of points in the scene are re-projected assuming an IO of 62.5 mm. For an IA of 62.5 mm the re-projected points would correspond with the real world points. Distances between real world points a and b and between c and d are shown for IA of both 35mm and 68.21mm. All points would be perceived in the same direction and the x-axis offset between the groups of points is for illustrative clarity only. The x-axis is not labelled as it is inconsequential since figure portrays the how the objects in space will be positioned in depth based on IA.

2.2.2 Motion in Depth Cues in the Visual System

As noted above, with variation in IA the geometry of S3D predicts that there are non-linearities in the relationship between distances in the captured scene and perceived stereoscopic depth. These distortions of space in the 3D scene have implications for perceived velocity (first derivative of depth with respect to time) and perceived acceleration (the second derivative of perceived depth with respect to time) of an object passing through that space. That is, the object moving at a constant speed should be perceived to travel at a higher velocity when the space is expanded and at a lower velocity when the space is compressed. As can be seen in Figure 2.1, a small IA predicts an overall compression of stereoscopic space and thus lower perceived velocity; conversely a large IA predicts increased perceived velocity.

Figure 2.1 also demonstrates the compression and expansion of space is not uniform (Diner, 1991). For a small IA the difference in predicted depth for a given real depth interval becomes increasingly larger as it is brought nearer (a-b < c-d); for a large IA the converse is true (a-b > c-d). By definition an object travelling at a constant velocity covers a constant real world distance per unit time. If the object approaches then this predicts that the object will appear to cover an increasing interval of stereoscopic space per unit time when captured with a small IA and hence appear to accelerate (Benzeroual, Allison, et al., 2011). Conversely, an approaching object should appear to decelerate when captured with a large IA. As far as we know, it has not been determined if viewers do in fact perceive these accelerations and decelerations.

2.3 Experiment 1 - Effects of Interaxial on Perceived Acceleration

Experiment 1 investigates how an observer's perception of the acceleration or deceleration of an object moving through stereoscopic space is affected by stereoscopic camera rig parameters, specifically the camera IA. Analysis of 3D geometry shows that the amount of depth predicted from stereopsis can vary non-linearly with the depth rendered/captured in the perspective images. It is an empirical question whether this dissociation will give rise to perceptual artefacts when viewing an object moving through that space.

As outlined in the Introduction, and illustrated in Figure 2.1, when image size (magnification), viewing distance and ZPS are kept constant, a change in the IA compresses or expands the predicted stereoscopic depth of the scene. It stands to reason that in a compressed space the object's predicted velocity will be less than in an expanded space since it traverses the two distances in the same interval of time. Thus, we predict that when an object moves through a scene containing these geometric distortions at a constant velocity, it should produce a non-linear velocity profile if it is seen according to the parallax, maintain a constant velocity if it is seen according to the monocular cues, or be perceived somewhere in between if monocular and binocular cues are combined. We evaluated this by varying the acceleration of an object moving through a computer-generated scene. We determined the acceleration at which the percept changed from acceleration to deceleration and vice versa to determine how these 'null' points shift with IA. The prediction is that if stereoscopic distortion introduces a bias toward acceleration or deceleration then the object will need to be ‘physically’ decelerated or accelerated (or virtually in the case of an animation) to counteract this bias.

2.3.1 Method

2.3.1.1 Subjects

Participants (n=9) ranged in age from 19 to 35, had normal or corrected to normal acuity, and good stereoscopic vision (assessed using the Randot Stereotest). Six of the observers were female, three were male, and all were paid for their participation. The experiments were approved by the York University Research Ethics Board and participants provided their informed consent.

2.3.1.2 Apparatus and Stimuli

Stimuli were created using Houdini 11.1 software (Side Effects Software, Toronto) and consisted of a 3.5s movie clip of a green ball rolling on the ground plane from a simulated distance of 9 m towards the observer to a distance of 2 m. To investigate how the presence of monocular cues affects perceived acceleration, the stimulus was presented under two conditions, with a textured (black and white checker) and non-textured (uniform reddish) ground plane. Figure 2.2 and Figure 2.3 show the stimuli with textured and uniform ground plane.

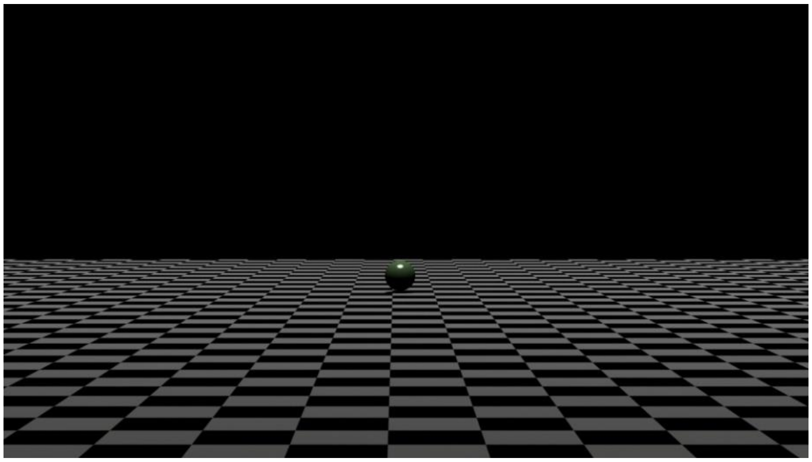

Figure 2-2

Figure 2.2 shows stimulus with a textured ground plane.

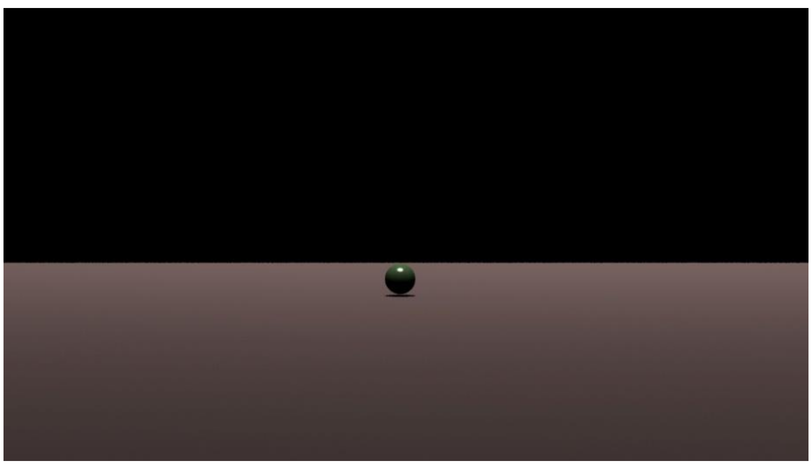

Figure 2-3

Figure 2.3 shows stimulus with a uniform ground plane

The ZPS was fixed at 3 m from the camera by horizontally shifting the image, and the focal length of the cameras was 88.23 mm. Four IAs were used to generate the stimuli: 35, 57.4, 65.7, and 68.21 mm. We generated predictions for the perception of acceleration from the geometry of binocular disparity assuming an object moving at constant velocity. As the predicted distortion of space is non-linear, each IA has its own predicted acceleration value: 35 mm = 34.62 cm/s^2, 57.4 mm = 16.65 cm/s^2, 65.7 mm = -16.65 cm/s^2 and 68.21 mm = -34.62 cm/s^2.

A series of clips were rendered with the ball decelerating or accelerating during its motion through the virtual space. For each combination of background and IA, thirteen clips were generated with 'physical' acceleration of the stimulus (through the virtual world as opposed to induced by stereoscopic distortion) ranging from -51.84 cm/s^2 to 51.84 cm/s^2 in steps of 8.64 cm/s^2. The average velocity was fixed at a mean velocity of 200 cm/s so that the length of the clips was the same and the ball rolled out of the frame at the same point. Thus, the initial velocities for each sequence differed depending on its acceleration value. The initial velocity values were 290, 275.5, 260.4, 245.3, 230.2, 215.0, 199.9, 184.8, 169.7, 154.6, 139.4, 124.3, and 109.2 cm/sec for the thirteen trials.

2.3.1.3 Procedure

We used a classical psychophysical technique known as the method of limits to estimate the point at which the participant saw the ball as neither accelerating nor decelerating (the 'null' point). For each estimate of this point for each condition, the stimulus was shown repeatedly starting with it obviously accelerating or decelerating. For the ascending runs the stimulus initially appeared to decelerate and on each subsequent presentation the acceleration was increased in steps of 8.64 cm/s^2. After each presentation the subject indicated verbally whether the ball appeared to accelerate or decelerate. The ascending run was terminated when the participant first responded ‘accelerating’ and the velocity on that trial was recorded. The reverse procedure was used for descending runs starting with an obviously accelerating stimulus, and decreasing the velocity until deceleration was reported.

For each ground plane and IA combination, four measures were made of the transition points between perceived acceleration and deceleration (two ascending and two descending trials per condition) for a total of 16 ascending and 16 descending runs. Given that ascending and descending series often result in different threshold biases, it is necessary to average equal numbers of ascending and descending runs to obtain threshold estimate (Ehrenstein & Ehrenstein, 1999). To familiarize the participants with the task and the stimuli, they were shown a sample series of both accelerating and decelerating trials prior to testing, and when needed, participants were given a short break at the half way point.

2.3.2 Results

Figure 2.4 shows the mean transition points (average of accelerating and decelerating runs) as a function of IA for the uniform and textured ground plane conditions. Also shown here are the geometrically predicted null values (white bars). The predicted null points are opposite in sign to the predicted bias because acceleration of the opposite sign must be ‘added’ to cancel this bias.

We predicted that the transition points would vary as a function of IA (see white bars in Figure 4). It is clear that these predictions were not born out. A repeated-measures ANOVA (with Greenhouse-Geiser corrections) confirmed that while there was a significant main effect of texture ($F(1,8) = 11.828$, $p=0.009$) there was no significant effect of IA ($F(3,24) = 0.853$, $p=0.465$). The interaction between IA and background did not reach significance ($F(3,24)=4.190$, $p=0.057$).

Figure 2.4 also shows that the presence or absence of texture in the ground plane influenced the acceleration/deceleration transition points. In the textured ground plane condition there was a bias towards seeing the stimuli decelerate (acceleration was required to null the bias). The converse was true in the uniform ground plane condition. This bias was consistent across IA in each background condition.

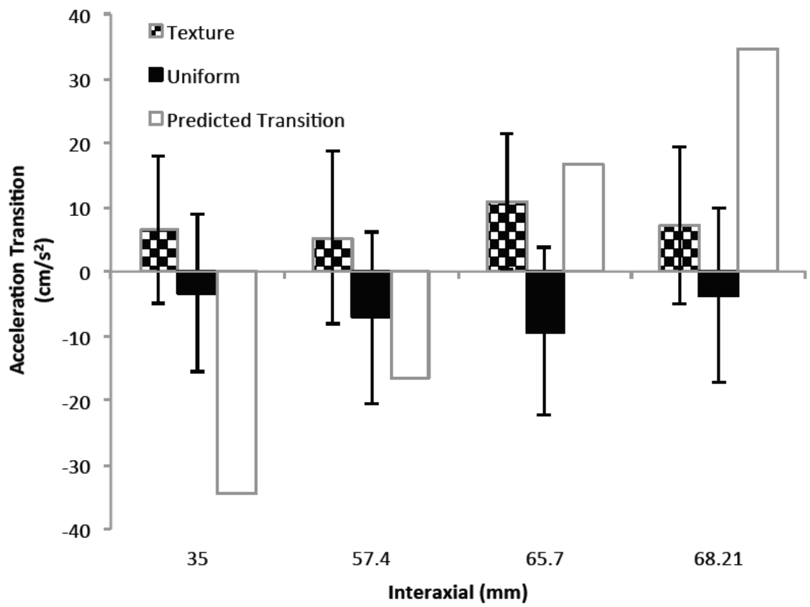

Figure 2-4

Figure 2.4 shows perceived mean transition points (average of ascending and descending runs) as a function of IA for the Uniform (black) and Textured (checked) background conditions. The error bars represent 95% confidence intervals and the white bars show the geometric predictions for each IA.

In summary, the results of Experiment 1 show that although observers' perception of the transition between perceived acceleration and deceleration differed significantly between the uniform and textured conditions, there was no consistent effect of IA in either condition.

2.4 Experiment 2 – Monocular and binocular contributions to acceleration bias

In Experiment 1, we evaluated the effect of predicted distortions of stereoscopic space on an object moving through that space. While there was no effect of the primary stereoscopic manipulation (IA) there was an effect of the monocular cue (ground plane texture). However, the observed bias could be solely due to the monocular perspective cue or to associated changes in the strength of the stereoscopic cue. That is, in the textured background condition, there are vertical edges throughout the lower field which may have been used by the stereoscopic system to specify the slant of the ground plane. Without these edges (in the uniform ground plane condition) the perceived slope of the ground plane may have been much weaker. Therefore, in Experiment 2 we isolate the monocular depth cues (e.g. perspective, looming, distance from horizon) to evaluate their contribution to the acceleration/deceleration transitions reported in our first experiment.

2.4.1 Method

2.4.1.1 Subjects and Apparatus

Twelve participants (ages 19-35) with normal or corrected to normal acuity were recruited. Of these seven were females and five were males. Stereopsis was assessed using the Randot Stereo test. The apparatus, camera parameters and viewing arrangement were the same as that described for Experiment 1.

2.4.1.2 Procedure

Observers viewed a moving ball sequence (as described in Experiment 1) either monocularly or binocularly. To equate as much as possible the monocular perspective views we used the smallest IA (35mm) employed in Experiment 1. The binocular trials were identical to those described in Experiment 1 while on monocular trials the participants wore an eye patch over their non-preferred eye. The method of limits was also used here to estimate the acceleration / decelaration transition points. Two ascending and two descending runs were performed for each combination of viewing condition (monocular/ binocular) and ground plane (textured / uniform) for a total of 16 runs per observer.

2.4.2 Results

Figure 2.5 shows the averaged perceived transition points in the uniform and textured ground plane conditions with their associated confidence intervals. The trends observed in Figure 2.5 are consistent with Experiment 1. While it appeared that there was a slight bias toward deceleration for textured backgrounds (nulled by an accelerating transition point) and acceleration for uniform backgrounds in the monocular viewing, the biases were not significant in any condition as indicated by confidence intervals in Figure 2.5.

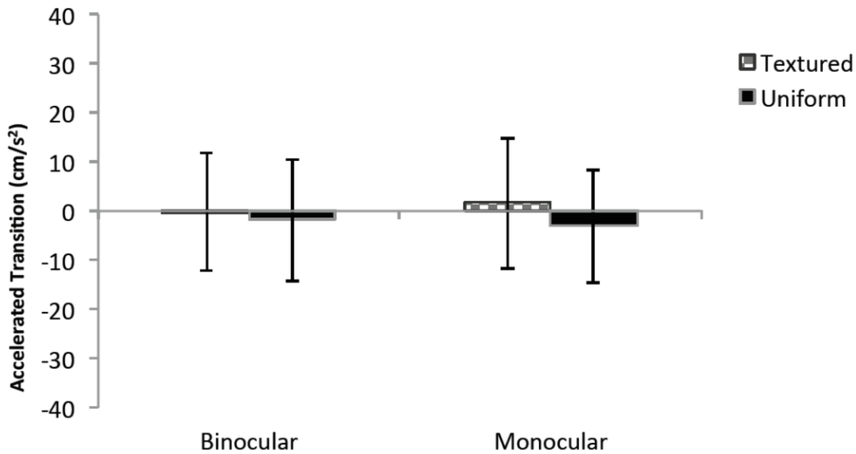

Figure 2-5

Figure 2.5 shows the average of perceived transition points of ascending and descending runs in monocular and binocular viewing conditions with textured and uniform ground planes. The error bars indicate the 95% confidence intervals for each data set.

Taken together, the results of Experiments 1 and 2 suggest that the distortions of stereoscopic space predicted from binocular viewing geometry do not influence the change in perceived velocity of an object moving through that space. While there appears to be some indication of an effect of the presence of monocular depth cues in the uniform ground plane condition, it is weak and in the same direction as that seen in the stereoscopic viewing condition. Thus, the results of Experiment 2 suggest that the availability of monocular cues was not responsible for observers' inability to perceive differences in acceleration transition points resulting from changes IA in Experiment 1.

2.5 Discussion

Experiment 1 shows that although observers' perception of the transition between perceived acceleration and deceleration was significantly affected by the presence or absence of a textured ground plane, there was no consistent effect of IA. The effect of texture may have been due to enhanced screen parallax, additional monocular depth cues, or some combination of these factors. Experiment 2 was conducted to examine how presentation of monocular depth cues in isolation contributed to the acceleration / deceleration transitions reported in Experiment 1. However, in Experiment 2 the bias was not significant and did not differ between monocular and binocular viewing conditions. Recall that the lack of bias in the binocular condition is consistent with the results of Experiment 1 as we used the smallest IA of 35 mm (though the bias is somewhat smaller in Experiment 2). Experiment 1 showed that strengthening the monocular and binocular depth signal changed the overall bias towards deceleration, suggesting that these cues do influence the perceived change in velocity. But this change occurred equally for all IAs, and the observed biases show no evidence of the predicted pattern, either in size or direction.

At the outset we outlined some relatively simple predictions regarding geometrically-defined distortions of S3D space and their possible effect on the perception of the motion of an object moving through that space (see Figure 2.1). We anticipated that the predicted depth distortions might influence the perceived acceleration / deceleration of an object moving through S3D space. However, it appears that such effects either do not occur, or are not captured using our methodology.

It may be that the predicted acceleration is simply too small to be detected or that the stereoscopic visual system is simply insensitive to acceleration. Humans have a notable lack of sensitivity to acceleration of objects across the retina. For example Gottsdanker et al. (1961) found that thresholds for discriminating acceleration in the frontal plane when both viewing duration and acceleration were varied were better explained by the difference between final and initial velocity than by acceleration. Similarly, Brouwer et al. (2002) reported that acceleration thresholds for both acceleration matching and discrimination tasks varied with both presentation time and mean velocity. However, expressed as a percentage change in velocity, the thresholds were constant at about 25%, independent of both presentation time and velocity. Thus, they also concluded that judgements of acceleration were made by comparing the initial and final velocity rather than directly detecting the acceleration. Practically, for our experiments it may not matter if perceived acceleration for motion in depth is sensed directly or through detection of change in velocity as both predict a perceptual artefact if the change is large enough.

Generally, subjects seem relatively insensitive to acceleration in the frontal plane (i.e. the plane of the screen) or to angular acceleration on the retina—thresholds for detecting the change in speed are typically 25% or more (Brouwer et al., 2002) which are significantly larger than the 5% typically reported for velocity discrimination(Snowden & Braddick, 1991). Less is known about sensitivity to acceleration in depth. For an isolated disk stimulus observers are reportedly very poor at discriminating simulated acceleration based on looming alone and are unable to discriminate a stimulus undergoing a 2 m/s^2 acceleration from a subsequently presented constant velocity stimulus when the stimuli are presented for 1s (Trewhella, Edwards, & Ibbotson, 2003). To date, to our knowledge, sensitivity to monocular acceleration in depth in a full cue environment has not been investigated. It is also not clear how sensitive viewers are to acceleration in stereoscopic motion in depth. However, it has been reported that simulated acceleration had no influence on either monocular or binocular time to contact judgements (López-moliner, Maiche, & Estaún, 2003). While this suggests that the binocular visual system is not sensitive to acceleration it should be noted that their stimulus was a single object with looming and changing disparity; results may differ if an object is presented in an environment with both monocular and binocular cues to relative motion.

Past research on velocity discrimination shows that unlike size and depth constancy, which scale image size and disparity according to distance, there is no evidence for velocity constancy. There appears to be no efficient means of compensating for depth or distance in the binocular mechanisms that are used to estimate the linear velocity of a moving object (McKee & Welch, 1989). Thus, even in a real world situation, given the variable relationship between changing disparity, changing depth, and the apparent lack of velocity constancy, it can be predicted that viewers should not be able to account for speed distortions of a moving object in depth. Consistent with this Rushton & Duke found that observers were unable to compensate for distance when judging velocity of motion in depth unless the stimulus was presented in a rich cue-laden environment (Rushton & Duke, 2009). If this is the case it may be that the insensitivity to stereoscopic acceleration distortions seen here may reflect adaptive mechanisms which downplay these distortions and create a bias toward perceiving constant velocity motion.

The presence of monocular cues to depth might also be expected to reduce the effects of IA-produced velocity distortions. The lack of a measurable effect of IA suggests that if cue conflict was the cause then the monocular cues were strong enough to result in cue dominance (Bülthoff & Mallot, 1988). In Experiment 1 since there was no difference in the effects of IA between the textured and uniform backgrounds, the looming and height in the field cues would have to be sufficient for this dominance. There is an extensive literature concerning the combination of monocular and binocular cues to depth (Howard & Rogers, 2002) and considerable evidence of interaction between these cues to produce the final percept (Hildreth & Royden, 2011; Young, Landy, & Maloney, 1993). Numerous studies report that there are individual biases towards perception of depth from one cue over another, for instance in the case of texture and disparity (Buckley & Frisby, 1993). There is also evidence of suprathreshold monocular depth cues such as motion parallax, looming and perspective dominating stereopsis. However typically these strong biases are observed when the disparity signal is weakened, for instance by using absolute disparity (Howard, 2008) or when strong depth cues are placed in conflict (Allison & Howard, 2000; Girshick & Banks, 2009). In our case, the predicted distortions are relatively large, and the parallax signal is above threshold, but it remains possible that the monocular cues were sufficient to override the parallax-based distortions.

Stereopsis can also interact with monocular cues in other ways. For example, the consistent disparity of the ball relative to the environment during the motion reinforces the relative position of the ball in the environment and supports the perception of motion in depth in the perspective image. As discussed in the Introduction, disparity needs to be scaled according to distance in order to recover metric depth. For many tasks in everyday life such scaling is not required and it has been argued that humans use simpler, non-metric relief transformations that preserve depth order and relief structure (Garding, Porrill, Frisby, & Mayhew, 1996). This relief transformation captures depth up to a scaling factor and would allow for an interpretation of the disparity to best match the depth from perspective, which could be thought of as a form of normalization.

It is possible that the absence of an effect of IA (and therefore stereopsis) seen here is due to a form of rescaling of perceptual space. Normalization processes operate in sometimes unexpected ways in stereoscopic media. In some cases, spatial distortions from nonlinear mapping of real world space to stereoscopic space can lead to an unnatural feel of the scene with artefacts such as the puppet theatre effect, cardboard effect, miniaturization and gigantism. Although varying IA can counter some of the artefacts in S3D content, at the extremes this variable can also cause unwanted distortions such as miniaturization (objects appear toy-like) and gigantism (objects appear over-sized). Typically, these phenomena cause the scene to appear as it would if it was a scaled version viewed from a nearer or further distance (Masaoka et al., 2006; Yamanoue et al., 2006) with the proportions of objects preserved but not their scale. If an object moved through such a normalized scene it might be expected that its velocity might also be normalized.

Our results are not consistent the proposal that observers should see distortions in motion in depth when IA is varied, and to our knowledge, this is the first evidence that the human visual system has this insensitivity. As outlined above, there are a number of likely causes of this insensitivity, and more research is needed to determine their relative contribution. In an effort to determine what the visual system does under typical viewing conditions, we used naturalistic motion and scene disparities (i.e. we did not distort the depth map). Our work shows that at least under such conditions, 3D content producers have flexibility in their choice of IAs without compromising the perception of an object's acceleration through stereoscopic space.

2.6 Conclusions

As outlined here IA, ZPS, viewing distance and a number of other camera parameters affect how distances in the real scene are mapped onto S3D space. Thus, modification of these variables has serious implications for perceived depth in S3D content. Recent advances in rendering software have made it possible to apply non-linear disparity mapping techniques to modify depth composition in previously captured content. The resulting content will have parallax-warped regions that do not obey the continuous geometry of visual space. Instead there may be abrupt transitions in stereoscopically defined depth between neighbouring regions in a scene. While in static scenes this distortion may not be noticeable, the motion of an object passing through these regions may be visibly distorted, drawing unwanted attention to that part of the scene. The experiments reported here investigate sensitivity to changes in velocity within regions that contain linear distortions that result naturally from binocular geometry. The next step will be to apply these methods to images containing non-linear disparity warping to determine if the observed insensitivity to S3D acceleration/deceleration is generalizable.

3 The Effects of Depth Warping on Perceived Acceleration in Stereoscopic Animation

3.1 Chapter Summary

Stereoscopic media produce the sensation of depth through differences between the images presented to the two eyes. These differences arise from binocular parallax which in turn is caused by the separation of the cameras used to capture the scene. Creators of stereoscopic media face the challenge of depicting compelling depth while restricting the amount of parallax to a comfortable range. To address this tradeoff, stereoscopic warping or depth adjustment algorithms are used in the post-production process to selectively increase or decrease the depth in specific regions. This process modifies the image's depth-to-parallax mapping to suit the desired parallax range. As the depth is adjusted using non-linear parallax re-mapping functions, the geometric stereoscopic space is distorted. In addition, the relative expansion or compression of stereoscopic space should theoretically affect the perceived acceleration of an object passing through that region. Here we evaluate this prediction and determine if stereoscopic warping affects viewers' perception of acceleration. Observers judged the perceived acceleration of an approaching object (a toy helicopter) moving in depth through a complex stereoscopic 3D scene. The helicopter flew at one of two altitudes, either ground level or camera level. For each altitude, stereoscopic animations were produced under three depth re-mapping conditions (i) compressive, (ii) expansive, and (iii) zero (no re-mapping) for a total of six test conditions.

We predicted that expansive depth re-mapping would produce a bias toward perceiving deceleration of the approaching helicopter, while compressive depth re-mapping would result in a bias toward seeing acceleration. However, there were no significant differences in the amount or direction of bias between the re-mapping conditions. We did find a significant effect of the helicopter altitude, such that there was little bias in acceleration judgements when the helicopter moved at ground level but a significant bias towards reporting acceleration when the helicopter moved at camera level. This result is consistent with the proposal that observers can make use of additional monocular (2D) cues in the ground level condition to improve their acceleration estimates. The lack of an effect of depth re-mapping suggests that viewers have considerable tolerance to depth distortions resulting from stereoscopic post-processing. These results have important implications for effective post-production and quality assurance for stereoscopic 3D content creation.

3.2 Introduction

Depth from disparity in a stereoscopic image reflects the layout of the scene but is also affected by a number of artistic, on-set production, post-production, and display parameters. A tool that filmmakers often use to determine stereoscopic parameters is the disparity range or parallax budget. This range is usually limited and carefully controlled to ensure comfortable viewing or to respect quality assurance standards. Thus, one of the most important stereoscopic 3D (S3D) filmmaking parameters is the interaxial distance between the two cameras, as it controls the mapping of the range of depths in the scene to the range of parallax in the image. This parallax is usually expressed in terms of the number of pixels of lateral shift between the images of the same object in the left and right camera view, or alternatively the size of this shift as a percentage of the image width. Assuming a 'typical' viewing geometry, the parallax budget controls the desired range of retinal disparities that will be presented to the user.

In native capture stereoscopic filming, the camera interaxial separation affects the parallax to depth relationship throughout the entire scene. With the single parameter of interaxial, parallax cannot be increased at one range of distances and simultaneously decreased or held constant at another. Once the content has been captured at a particular camera baseline, it is generally not possible to change the interaxial. Furthermore, during capture, the parallax range is typically optimized for a particular screen size; when the footage is subsequently displayed on a screen with a different size, the predicted depth experienced by the audience will change. For example, according to stereoscopic geometry, if a S3D image optimized for a theatre-sized screen is viewed on a small screen (at a nearer distance so it subtends the same visual angle) then geometry predicts a decrease in depth and a flattening of objects in the scene. Conversely, an image designed for a small screen can be uncomfortable to view when scaled up. Although S3D content producers have various capturing techniques, such as multi-view acquisition and multiple retakes, at their disposal to adapt the content to different viewing conditions, these methods can be computationally complex, time-consuming, and costly. Content producers can also employ labour intensive tools such as manual editing of disparities by compositing scenes from multiple stereo rigs with varying baselines.

To cope with these issues, S3D content producers can use stereoscopic warping or depth adjustment in the post-production phase. This process permits modification of the image's depth-to-parallax mapping to suit the desired parallax range. In addition, depth-to-disparity modelling is performed such that the parallax can be selectively retargeted for a defined region of interest in the image. This results in modified depth in specific parts of the 3D image leaving the depth in other regions unchanged. This kind of region-defined disparity manipulation is known as non-linear parallax mapping or non-linear disparity mapping by some researchers, and provides the content producers with a robust means to manipulate and control depth in isolated parts of the image for technical and/or artistic reasons (Feldmann, Schreer, Kauff, & Nachrichtentechnik, 2003; Jones, Lee, Holliman, & Ezra, 2001; Kellnhofer, Ritschel, Myszkowski, & Seidel, 2013; Lang et al., 2010; Niu, Feng, & Liu, 2012).

Non-linear depth adjustment can be beneficial when a large range of distances need to be portrayed in a scene that also contains a volumetric object in the foreground. If this scene is shot with a restricted parallax budget, then all depths within the scene will be compressed, and the foreground object may appear significantly compressed and flattened relative to its width. If the depth budget is expanded, to give the foreground object(s) more volume, the director risks introducing too much parallax in the background which may introduce divergence. Region-specific disparity manipulation through non-linear disparity mapping provides a solution as it permits enhancement of the depth of the foreground, while leaving the depth in the background unchanged (Jones et al., 2001).

The non-linear parallax mapping process involves several steps including: (i) generation of parallax or depth maps, (ii) adjustment of the function between z_{in} (distance from the camera in the original stereo pairs) and z_{out} (the geometrically predicted distance from the viewer of the display), and (iii) creation of a new stereoscopic image based on the depth maps and the new z_{out} function. The key element is that the user or 3D content producer can specify the in-out distance or parallax maps based on the desired creative and technical effects. It is a common practice that once the left view has been acquired, the parallax manipulations are performed on the right view based on the nearest and farthest distances, desired parallaxes at these points, and the parallax remapping function between these end points. As the depth is adjusted using non-linear parallax re-mapping functions, the geometrically predicted stereoscopic space is distorted. It no longer adheres to the stereoscopic geometry of the scene that was captured (even if presented orthostereoscopically) or to the monocular depth cues in the perspective images.

Typically, the input-output parallax mapping functions are monotonic and smooth to avoid reversals or discontinuities in the relationship between parallax and depth (discontinuities can be used if no objects appear at or near the distance of the discontinuity). As noted in the figures that follow, unless they are linear, these monotonic parallax remapping functions produce stereoscopic distortions; depths between the different points of the scene are compressed or expanded depending on their distance. In expansive parallax re-mapping functions, the parallax between objects that are near the camera will be smaller than in the original stereoscopic images, while the parallax will be larger than in the original at far distances from the camera. This predicts a perceived expansion of stereoscopic space with increasing distance from the camera. Conversely, with a compressive parallax re-mapping function, compression of space should increase as distance from the camera increases.

As well as affecting static depth intervals, this compression or expansion of stereoscopic space should affect the perceived motion of an object traversing the space. That is, if an object moves in depth relative to the camera at constant velocity in the scene, but the mapping from the original to transformed stereoscopic is compressive or expansive, then the object's apparent velocity should change over time. If the object moves in the direction in which the space expands then it should appear to accelerate. Conversely, it should appear to decelerate if it moves in the opposite direction.

3.3 Methods

3.3.1 Participants

Participants (n=7) ranged in age from 19 to 35, had normal or corrected to normal visual acuity, and good stereoscopic vision (at least 40 arcseconds assessed using the Randot Stereotest). Three of the observers were female, four were male, and all were paid for their participation. The experiments were approved by the York University Research Ethics Board and participants provided their informed consent.

3.3.2 Stimuli

The scene was constructed in Maya (Autodesk, version 2012) to simulate a realistic animation clip from an S3D movie complete with multiple monocular and binocular depth cues. Each graphic unit in the modelled scene was equivalent to a centimeter in 3D space and thus we will report all dimensions in cm (see below for the effects of image scaling and stereoscopic presentation). The frame rate of the stereoscopic presentation was 24 fps and we will report all durations, speeds and accelerations in seconds based on this frame rate.

The shot consisted of a toy helicopter (38.51 cm x 12.85 cm) flying toward the camera across a school gym (604.52 cm by 998.98 cm). The gym included a stage with curtains in the background, multiple rows and columns of chairs placed in the middle of the gym, as well as various other objects placed throughout the space (Figure 3.1). The only moving object in the scene was the helicopter, which sported a rotor that turned at a constant speed throughout the clip (so as not to provide an extraneous cue for the acceleration or deceleration of the helicopter).

In the z-space (depth direction) of the scene, the helicopter moved from a distance of 650 cm to approximately 150 cm. Because the camera was located at z=25 cm this corresponded to a range of 625 cm to 125 cm relative to the camera. The z-axis motion path specified in Maya was a straight line approach trajectory traversing 500 cm. The helicopter covered this distance in 5 s (120 frames) for a simulated average speed of 100 cm/s. This average speed was constant across animations although acceleration, and thus the start and end speed varied.

Figure 3-1

Figure 3.1 shows the helicopter shot / gym scene used in the experiments. The helicopter is approaching at floor level in this image.

Figure 3-2

Figure 3.2 shows the helicopter approaching at eye level.

The position of an object relative to the ground plane is a strong monocular depth cue (called ‘height in the field’). Further, changing the altitude of the object will influence the availability, and influence of, monocular depth cues. Thus, we evaluated whether the height of the helicopter above the ground had an impact on the viewer’s perception of acceleration or deceleration. To do so we compared two ‘altitude’ conditions: one where the simulated helicopter flew just above the ground and one where it approached at camera height (43.7 cm). In the latter case the helicopter remained at the same vertical position in the image as it approached, making the height in the field cues uninformative (Figure 3.2).

In all the conditions, the start position of the helicopter was at 625 cm from the camera and the end position was at 125 cm from the camera in the world z-coordinates (i.e., separation in depth along the floor). However, the camera was pitched downward (-3.34º) in the floor level condition for better framing. This resulted in slight differences in distances along the camera z-direction of the start and end positions of the helicopter between the two conditions That is, they were slightly further from the camera (+4 cm) in the ground level condition than in the camera level condition.

Stereoscopic animations were rendered as separate left and right eye views. It was critical to maintain constant parallax at the start and end position of the helicopter throughout all the test conditions. In these experiments the effective unwarped stereoscopic camera rig had parallel cameras with focal length of 35 mm (35-mm equivalent, 54º horizontal field of view) and an interaxial separation of 1.98 cm. The rig was converged (using standard off-axis/skew frustrum stereoscopic rendering) at 125 cm from the camera so that the animation stopped as the nose of the helicopter reached the screen plane (i.e., had zero screen parallax). Based on these parameters, the parallax at the start position for the ground level was 25.92 pixels and for the eye-level condition was 23.92 pixels across all test conditions. The end disparity of the front tip of the helicopter was zero in all conditions.

The non-linear parallax remapping was accomplished based on the rendered camera frames and associated depth maps using Fusion (Blackmagic Design, version 6.3). Based on the input images, and the desired parallax remapping, a new right camera image sequence was rendered. The disparities at the start and end position of the helicopter's travel were controlled and maintained in Fusion and were verified by comparison with onscreen disparity. In the stereoscopic tools in Fusion, a near and far point can be associated with a range of parallax values (parallax budget) and the desired interaxial distance computed. These values were set to match the stereoscopic rig parameters listed in the previous paragraph.

3.3.3 Non-linear parallax re-mapping

Three non-linear parallax re-mapping functions were used for this study. The parallaxes at both the start (z = 650) and end (z = 150) positions of the helicopter were constant across all conditions and the parallax at these points corresponded to the distance in the modelled scene and the stereoscopic rig parameters. To remap the depth of the selected scene, to the space between these end points (i.e. along the path traversed by the helicopter) the relationship between distance in the modelled scene and the parallax specified distance was mapped according to one of three quadratic functions. These functions describe the relationship between the original z position of the helicopter in the modelled scene (z_{in}) and the z position specified by the screen parallax after mapping (z_{out}) (Refer to Figure 3.3 below):

1. Linear: $z_{out} = z_{in}$ In the linear condition the stereoscopic parallax matches the perspective depth. The correspondence is appropriate for the camera baseline and this

parallax is not manipulated in the post-production. Note, however, that the stereoscopic camera baseline is smaller than the observer's eye separation so the predicted stereoscopic depth for the whole scene is somewhat compressed relative to the modelled scene. This stereoscopic distortion in the scene is the natural result of the camera set up and parameters and is typical for stereoscopic movies.

2. Compressive function: $z_{out} = \begin{cases} -0.00124z_{in}^2 + 1.992z_{in} - 120.9, & 150 \leq z_{in} \leq 650 \\ z_{in}, & \text{otherwise} \end{cases}$

   where stereoscopic depth intervals are compressed relative to the model at distances far from the camera. Given that the object moves towards the camera (z-values decrease) the object should appear to accelerate as it approaches (relative to the linear case). As the helicopter covers the same distance (end points are fixed) in the same time average simulated velocity remains 100 cm/s. However, the velocity changes with a simulated acceleration of 24.8 cm/s^2, starting from 38 cm/s at z = 650 and increasing to 162 cm/s at z = 150.

3. Expansive function: $z_{out} = \begin{cases} 0.00124z_{in}^2 + 0.008z_{in} + 120.9, & 150 \leq z_{in} \leq 650 \\ z_{in}, & \text{otherwise} \end{cases}$

   where stereoscopic depth intervals are expanded relative to the model at distances far from the camera. Since the object moves towards the camera it should appear to decelerate (relative to the linear case). The simulated deceleration is 24.8 cm/s^2, starting from 162 cm/s at z = 650 and decreasing to 38 cm/s at z = 150.

Care was taken so that the rate of change of depth was either strictly increasing or decreasing over the manipulated trajectory (z-values between 650 cm and 150cm). In Fusion, disparity manipulation occurs at the nodal level, in which a node is created to specify how the z-value is manipulated. Thus, the three z_{in} versus z_{out} relations were generated over the range of relevant z values (Figure 3.3).

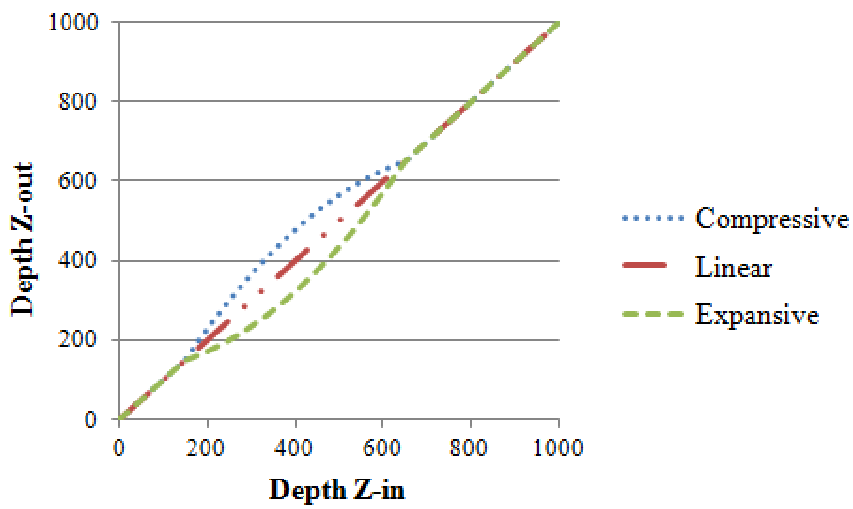

Figure 3-3

Figure 3.3 shows Non-linear parallax remapping function. Parallax for z_{in} values (x-axis) were remapped to parallax appropriate to a different z_{out} (y-axis).

3.3.4 Stereoscopic Display and Viewing

The stimuli were displayed on 54-inch 1080p Panasonic Viera Series TC-P54VT25 3DTV plasma television with dimensions of 119.8 (W) and 67.3 (H) cm and pixel resolution of 1920 (W) and 1080 (H). The stereoscopic images were viewed through active shutter glasses provided with the television. The position of the viewer relative to the screen was adjusted to obtain a horizontal viewing angle of 36 degrees (1.8m from screen). The viewer sat stationary throughout the session.

The camera interaxial separation was smaller than the viewer's interocular distance (which averaged 5.81 cm as measured with a Reichert, Digital PD Meter). Additionally, the CG camera field of view was wider than the display visual angle. These factors are typical of filmed stereoscopic content in commercial productions where lens and interaxial choices are made for reasons of artistic preference and visual comfort rather than to match a specific viewer and viewing position (which is impossible in a theatre setting).

3.3.5 Procedure

The method of constant stimuli was used to assess the perceived acceleration/deceleration of the helicopter under each of the disparity mapping and altitude conditions. For each condition, stimuli with 11 levels of acceleration were presented to cover a range of positive and negative accelerations. A set of pilot experiments determined an appropriate range of accelerations needed to adequately sample the psychometric functions (Table 3.1). On each trial, the observer viewed one helicopter trajectory, and indicated verbally whether the helicopter appeared to be accelerating or decelerating. Each stimulus condition was presented 20 times over 10 sessions (2 repeats per session). For each subject, the proportion of accelerating responses at each stimulus level was tabulated and psychometric functions were fit using a Weibull function. From each psychometric function we estimated the point of subjective equality (PSE) for each mapping function. The PSE in the present case is the stimulus level at which the participant is equally like to report that the helicopter was accelerating or decelerating (the 50% point on the psychometric function). Bias predictions were calculated using each observer's interocular distance. If the re-mapping biases the apparent acceleration in one direction, then the PSE should be shifted in the opposite direction since acceleration in the opposite direction is needed to null this bias.

Table 3.1

Acceleration ($cm/frame^2$)	**Acceleration (cm/s^2)**	**Initial velocity (cm/s)**	**Final velocity (cm/s)**
0.05	28.80	30.0	172.8
0.04	23.04	44.3	158.6
0.03	17.28	58.6	144.3
0.02	11.52	72.9	130.0
0.01	5.76	87.2	115.7
0.00	0.00	101.4	101.4
-0.01	-5.76	116.4	87.8
-0.02	-11.52	130.7	73.5
-0.03	-17.28	144.9	59.3
-0.04	-23.04	159.2	45.0
-0.05	-28.80	173.5	30.7

Table 3.1. Acceleration levels used to measure the effects of non-linear parallax remapping.

3.4 Results and Discussion

Figure 3.4 shows sample psychometric functions for one observer for the linear, compressive, and expansive non-linear parallax re-mapping conditions. As expected, as acceleration increases from negative to positive, the proportion of positive acceleration responses increases. The observer in Figure 3.4 demonstrates bias that is dependent on the remapping function. Once again, it should be noted that the predicted PSE (50% point on the fitted curve) is opposite in direction to the bias because oppositely directed acceleration will be needed to cancel the bias. For example, in Figure 3.4, the observer was biased to perceive the stimulus as accelerating for the compressive function; however, because the PSE reflects the amount of acceleration needed to cancel the positive bias, the PSE value is negative.

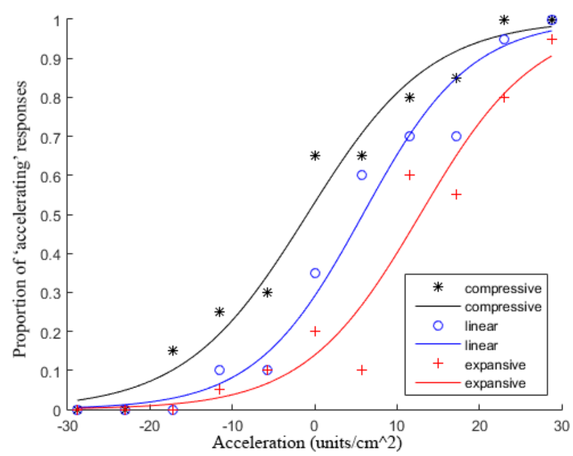

Figure 3-4

Figure 3.4 shows sample psychometric functions (fit with Weibull functions) for one subject for each mapping condition (compressive, linear and expansive) of the ground condition.

The average PSEs for all participants are shown in Figure 4.55 as a function of the non-linear parallax remapping condition and altitude. This figure also depicts the geometrically predicted PSE values based on the stereoscopic geometry.

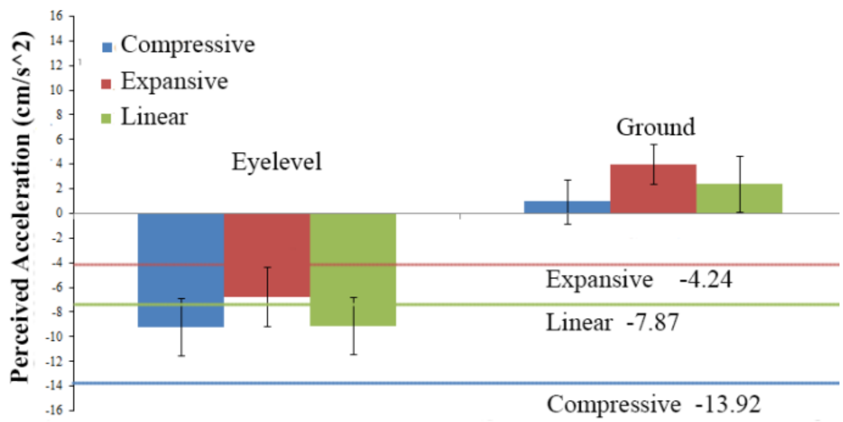

Figure 3-5

Figure 3.5 shows average PSE values for eye (camera) level and ground level conditions. The lines show predicted PSE values for each remapping curve with compressive at -13.9, linear at -7.9, and expansive at -4.2 cm/s^2

We predicted that the transition points would vary as a function of non-linear parallax remapping algorithm. Furthermore, when the stereoscopic capture parameters differ from the viewing parameters (as in this study and in most filmed content) geometry predicts a transformation between the geometry of the modeled scene and its appearance in the perceptual display space (Allison, 2004; Woods, 1993). In the current scenario the differences in baseline and magnification predict a nonlinear compression of space relative to the original scene. As a result of the nonlinearity, we would predict—based on parallax alone—that an object moving at constant velocity in depth would appear to be accelerating at 7.9 cm/s^2. The acceleration predictions for our compressive and expansive re-mappings would also be subject to this compression and hence biased toward acceleration with predicted accelerations of 13.9 and 4.2 cm/s^2, respectively.

It is clear that these predictions were not borne out. A repeated-measures ANOVA confirmed that while there was a significant main effect of altitude ($F(1,35)=37.438$, $p=5.39e\text{-}07$), there was no effect of depth warping on perceived acceleration ($F(2,35)=0.857$, $p=0.433$). The interaction between depth warping and the altitude of the helicopter was also not significant ($F(2,35)=0.046$, $p=0.955$).

The impact of helicopter altitude on perceived acceleration is evident in Figure 3.5. That is, in the ground level condition there was little bias, while in the camera level condition, there was a significant bias toward seeing the helicopter accelerate (deceleration was required to null the bias). This altitude-dependent bias was consistent across the non-linear parallax re-mapping conditions.

In summary, the results of the experiment show that although participants' perception of the transition between perceived acceleration and deceleration differed significantly as a function of helicopter altitude, there was no consistent effect of stereoscopic depth warping at either altitude.

The S3D scene contained monocular cues such as looming, perspective, and distance from the horizon in addition to the screen parallax. To isolate the effect of the monocular cues and their influence on perceived motion through the scene we replicated the main experiment but with monocular viewing of the linear mapping stimulus (with non-dominant eye patched) in three participants. We found no significant differences in PSE between eye level and ground condition when viewed monocularly. Two participants displayed bias towards seeing the helicopter accelerate, while the other was biased toward seeing deceleration. The bias toward acceleration was consistent with the results of the main experiment, thus confirming that monocular cues may have played a role in the perception of acceleration in the first experiment.

3.5 Discussion

The three main cues to motion-in-depth that have been studied in the literature are target vergence (absolute screen parallax of the helicopter in our study), relative disparity (difference in parallax, or relative parallax, between the helicopter and other stationary features), and changing image size / looming (here looming of the helicopter). One of the earliest studies done on this topic was by Regan and Beverly in 1979. Their experiments investigated the rate of change in disparity that was needed to null perceived motion in depth due to changing target size (Regan & Beverley, 1979). They determined that the contribution of disparity to motion in depth perception becomes more pronounced when either velocity or presentation times are increased. Subsequent studies have investigated the added role of target vergence, in addition to relative parallax and looming in perception of motion in depth.

For instance, Brenner et al. (Brenner et al., 1996) presented targets with simulated motion in depth based on various combinations of changing absolute screen parallax, changing relative parallax (presence or absence of other stationary features), and looming. Their participants adjusted the lateral velocity of a subsequently presented probe to match the target's speed of motion in depth. Matched lateral velocity was always less than the simulated motion-in-depth velocity. Changing screen parallax (target vergence) alone did not produce motion in depth (see also Erkelens & Collewijn, 1985) but could influence the matches when other cues were also presented. This suggests that target vergence alone was insufficient to produce motion in depth but could modulate it. Compared to our stimuli, the monocular cues to distance in these displays were minimal. Nevertheless, looming produced the most robust motion in depth and, somewhat unexpectedly as it was ambiguous to absolute distance, image size affected perceived target distance. Changing the screen parallax of the target, relative to static features, produced motion in depth although it was reduced relative to that experienced in the full cue or looming only conditions. Others have shown that the perceptual interpretation of the binocular cues depends on the strength of monocular looming cues. The latter can modulate the magnitude and even the direction of the perceived motion in depth (González et al., 2010; Howard et al., 2014).

Compared to these previous studies, in Experiment 1, rich monocular and binocular cues were available to specify the depth of the helicopter relative to the rest of the scene. Further, changing screen parallax was always accompanied by changing relative parallax. The absolute screen parallax of the helicopter at any time was nonlinearly mapped according to the depth remapping and thus the apparent motion due to target vergence should have manifested the predicted accelerations and decelerations.

Predictions based on relative parallax are more complicated. Relative to the unwarped distant background (or near foreground), the changing relative parallax of the helicopter was distorted in the same manner as the absolute parallax and thus we would predict acceleration and deceleration of the helicopter under parallax re-mapping. However, note that the entire scene was also subject to the same parallax re-mapping as the helicopter. Therefore, during the motion both the path of the helicopter and the space it moved through were stereoscopically distorted and thus the distortion of local relative parallax was small, particularly for the floor altitude condition. In fact, any object stereoscopically aligned with the helicopter at a given frame in the linear animation would also be aligned with helicopter at the corresponding frame in the re-mapped animations. Thus, when comparing the parallax (and predicted relative stereoscopic depth) of the helicopter over time relative to features at similar depth (say on the floor), the progress of the helicopter over the space should be identical regardless of the warping. In this case the distortion of the motion path of the helicopter might depend on the distortion (or lack of) of the depth in the scene itself.

We have some evidence that in the presence of strong perspective and other monocular cues such distortions may be minimal. Previous studies on the effect of interaxial separation on perception of stereoscopic depth have shown that observers have a surprising ability to tolerate a wide range of interaxial settings without significant distortion (Allison & Wilcox, 2015). If the apparent depth and shape of the scene were unperturbed by the re-mapping then we would predict that observer would not see the acceleration expected if the space was apparently distorted. McKee and Welch (McKee & Welch, 1989) have demonstrated that observers have poor ability to scale velocity for distance (velocity constancy) compared to size or depth. Therefore, even if stereoscopic space appeared distorted it may not have been evident in the motion profile of the helicopter.

Another important aspect of motion in depth studies related to our experiments is the observers' ability to perceive acceleration in moving objects. In an earlier study we assessed whether asymmetric distortion in the mapping of stereoscopic to real space as a result of varying interaxial has impact on the perceived acceleration of an object moving through that space under orthostereoscopic viewing (Laldin, Wilcox, Hylton, & Allison, 2012). Our data suggested that observers were able to discount distortions of stereoscopic space in interpreting object acceleration / deceleration (Laldin et al., 2012). Other previous work has suggested that observers are relatively insensitive to 2D or 3D acceleration. Gottsdanker et al.(1961) asked subjects to discriminate between accelerated and constant-velocity moving targets. They found that performance improved as mean velocity of the object increased, and presentation time decreased. They concluded, as did subsequent studies (Werkhoven, Snippe, & Toet, 1992), that rather than being perceived directly, acceleration was detected by comparing early and late velocities. Later studies have shown that observers often misperceive acceleration resulting in overestimation times of arrival time in time-to-contact experiments (López-moliner et al., 2003), although the bias was reduced when binocular information was present. More recent studies have demonstrated that observers have high thresholds and systematic biases when judging acceleration based on image looming (Mueller & Timney, 2014). This general lack of sensitivity to acceleration of moving objects may underlie the lack of an influence of depth remapping on apparent acceleration that we found in the present study. However, we note that observers were able to discriminate acceleration in our stimuli (Figure 4.4). The just noticeable differences in the psychometric functions were considerably smaller than the predicted effects of the nonlinear disparity remapping. Thus, participants should have been sensitive to the predicted effects of the remapping.

In contrast to the lack of an effect of parallax remapping we did find a significant effect of altitude. The small biases seen in the ground level condition suggests that observers were using ground plane cues in these judgements. As noted above, any factor that influences perceived distance should impact the perceived acceleration of an object traversing through that distance. The importance of the ground surface in determining the perceived distance of objects in 3D scenes was noted by Al-Hazen about 1000 years ago (Al-Hazen, 1989). Much later, Gibson (Gibson, 1950) highlighted the importance of the ground surface in perception of 3D scenes. One aspect is the so-called height in the visual field, which reflects the geometric relation between distance and height in the optic array. As the horizon lies at eye level, everything on the ground is projected onto the lower half of the optic array (Sedgwick, 1986). As the distance from the observer increases, so does the height in the optic array. More recent research by Bian and colleagues (Bian et al., 2005) has posited a ground dominance effect in determining the perceived relative distances of objects in 3D scenes. They found that when presented with vertical posts in optical contact with either the ground or the ceiling, the observers perceived the one attached to ground surface to be closer.

The increased accuracy of acceleration judgements at ground level compared to eye level that we found is similar to improvements in other judgements when ground contact information was available. For example, Philbeck & Loomis compared walking and verbal report as indicators of distance with angular elevation (eye-level vs. ground level) as a variable. They found that adding height in the field as a distance cue (moving the target from eye-level to ground level) improved the accuracy of walked distance, thus providing evidence of angular elevation as an effective cue to egocentric distance (Philbeck & Loomis, 1997).

Our study has important implications for both content production and quality assurance of stereoscopic 3D media. Content producers face the challenge of increasing depth and maintaining the parallax budget, thus depth warping has become a popular and effective tool. Our research shows that observers have significant tolerance to parallax re-mapping; they generally are not sensitive to the differences in motion that theoretically result from re-mapping functions.

4 General Discussion

The first experiment of this thesis investigated how an observer's perception of the acceleration or deceleration of an object moving through stereoscopic space was affected by the camera IA. Analysis of 3D geometry shows that the amount of depth predicted from stereoscopic viewing can vary non-linearly with the depth rendered/captured in the perspective images. Variations in IA predict distortion of space suggesting that an object moving at constant velocity should be perceived as accelerating by the viewer. In the first experiment, although it was predicted that the apparent acceleration would vary as a function of IA, these predictions were not borne out. The results showed that the presence or absence of texture in the ground plane influenced the acceleration/deceleration transition points. In the textured ground plane condition there was a bias towards seeing the stimuli decelerate. The converse was true in the uniform ground plane condition. This bias was consistent across IA in each background condition.

The second experiment also investigates perception of viewer acceleration in stereoscopic images; however, in the second experiment the stereoscopic space was warped to selectively increase or decrease the depth in specific regions. As the depth is adjusted using non-linear parallax re-mapping functions, the geometric stereoscopic space is distorted. As predicted in the first experiment, the relative expansion or compression of stereoscopic space should theoretically affect the apparent acceleration of an object passing through that region. However, in this case, the compression and expansion of space was deliberate rather than introduced as a consequence of camera IA. We predicted that the perceived acceleration would vary as a function of non-linear parallax remapping algorithm; specifically, as a result of the nonlinearity, we would predict that an object moving at constant velocity in depth would appear to be accelerating or decelerating based on the quadratic function. However, our results showed that the predictions were not confirmed. While there was a significant effect of altitude, there was no effect of depth warping on perceived acceleration. The interaction between depth warping and the altitude of the helicopter was also not significant.

Both studies suggest that viewers are tolerant of effects of stereoscopic distortions when perceiving accelerating or decelerating objects in a stereoscopic scene. Both studies inherently investigated interaction of various binocular and monocular cues to motion in depth including target vergence, looming, and relative disparity. Based on my results I concluded that in simple and complex stereoscopic scenes, these cues interact with each other in a way that makes viewers insensitive to stereoscopic distortion of motion in depth. Although several studies have noted lack of direct perception of acceleration in the frontal plane (as opposed to changes in the initial and final velocities), studies of the perception of acceleration in motion in depth have been relatively few. More recent studies have shown that observers have high thresholds when judging acceleration based on image looming. Furthermore, as outlined in the introduction, people can perceive motion in depth based on change-of-disparity (CD) signals, interocular velocity (IOVD) difference signals, or, finally, through specialized “dynamic disparity detectors”, that register changing disparity in binocularly matched features. Although there is evidence of stereomotion detectors, our studies show that these detectors are insensitive to perception of acceleration in stereoscopic scenes.

It is possible that under some conditions observers may be sensitive to distorted motion in depth. For example, past studies by have shown that increasing differences in initial and final velocity, as well as decreasing presentation times, helps with detection of acceleration in motion in depth (Gottsdanker et al., 1961). For future studies, it might be worthwhile to experiment with variable viewing times and increasing the distance travelled by the object in space. In the present studies, the ball in the first experiment travelled about 7 units (which equaled to 7 meters) and in the second experiment, the helicopter travelled 600 units (600 cm); the first experiment’s trial clip was 3.5 seconds long whereas the second experiment’s trial clip was 5 seconds long. Perhaps the distance covered and presentation times were not adequate to capture perception of acceleration in these particular scenes.

Our studies have implications for how stereoscopic media are created. Whereas geometry may predict perceptual effects based on stereoscopic imaging and display parameters, the predicted effects will not always be borne out. Although there are stereoscopic tools that use binocular geometry to set depth range, the use of these tools does not tell us what may be possible to create through stereoscopic media. This means that filmmakers have more leeway to use their creative license to stretch the boundaries of what geometry allows us to do to create interesting immersive experiences.

Bibliography

Al-Hazen, I. (1989). Book of Optics. In A. I. (Trans) Sabra (Ed.), *The Optics of Ibn-Haytham* (Volume I). London: University of London, Warburg Institute.

Allison, R., & Howard, I. P. (2011). Stereoscopic Motion in Depth. *Vision in 3D Environments*, (416), 1–37. https://doi.org/10.1017/CBO9780511736261.008

Allison, R. S. (2004). The Camera Convergence Problem Revisited. *SPIE Stereoscopic Displays and Virtual Reality Systems XI*, *5291*, 167–178. https://doi.org/10.1117/12.526278

Allison, R. S., Gillam, B. J., & Vecellio, E. (2009). Binocular depth discrimination and estimation beyond interaction space. *Journal of Vision*, *9*(1), 10.1-14.

Allison, R. S., & Howard, I. P. (2000). Temporal dependencies in resolving monocular and binocular cue conflict in slant perception. *Vision Research*, *40*(14), 1869–1886. https://doi.org/10.1016/S0042-6989(00)00034-1

Allison, R. S., & Wilcox, L. M. (2015). Perceptual Tolerance to Stereoscopic 3D Image Distortion. *ACM Transactions on Applied Perception (TAP)*, *12*(3), 10. https://doi.org/10.1145/2770875

Benzeroual, K., Allison, R. S., & Wilcox, L. M. (2012). 3D display size matters: Compensating for the perceptual effects of S3D display scaling. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 45–52. https://doi.org/10.1109/CVPRW.2012.6238907

Benzeroual, K., Allison, R., & Wilcox, L. M. (2011). Distortions of Space in Stereoscopic 3D Content. *SMPTE International Conference on Stereoscopic 3D for Media and Entertainment*, *1100*.

Benzeroual, K., Wilcox, L. M., Kazimi, A., & Allison, R. S. (2011). On the distinction between perceived & predicted depth in S3D films. *2011 International Conference on 3D Imaging, IC3D 2011*. https://doi.org/10.1109/IC3D.2011.6584389

Bercovitz, J. (1998). Image-side perspective and stereoscopy. *SPIE*, *3295*, 288–298. https://doi.org/10.1117/12.307152

Bian, Z., Braunstein, M. L., & Andersen, G. J. (2005). The ground dominance effect in the perception of 3-D layout. *Perception & Psychophysics*, *67*(5), 802–815. https://doi.org/10.3758/BF03193534

Bian, Z., Braunstein, M. L., & Andersen, G. J. (2006). The ground dominance effect in the perception of relative distance in 3-D scenes is mainly due to characteristics of the ground surface. *Perception & Psychophysics*, *68*(8), 1297–1309. https://doi.org/10.3758/BF03193729

Brenner, E., Van Den Berg, a. V., & Van Damme, W. J. (1996). Perceived motion in depth. *Vision Research*, *36*(5), 699–706. https://doi.org/10.1016/0042-6989(95)00146-8

Brouwer, A.-M., Brenner, E., & Smeets, J. B. J. (2002). Perception of acceleration with short presentation times: can acceleration be used in interception? *Perception & Psychophysics*, *64*(7), 1160–1168. https://doi.org/10.3758/BF03194764

Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research*, *33*(7), 919–933.

Bülthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: stereo and shading. *Journal of the Optical Society of America A*, *5*(10), 1749–1758.

Cumming, B. G., & Parker, A. J. (1994). Binocular mechanisms for detecting motion-in-depth. *Vision Research*, *34*(4), 483–495. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8303832

Czuba, T. B., Guillet, K., & Cormack, L. K. (2013). Three-dimensional motion aftereffects reveal distinct direction-selective mechanisms for binocular processing of motion, *11*(10), 1–32. https://doi.org/10.1167/11.10.18.Three-dimensional

Diner, D. B. (1991). Stereoscopic Camera and Viewing Systems with Undistorted Depth Presentation and Reduced or Eliminated Erroneous Acceleration and Deceleration Perceptions, or With Perception Produced or Enhanced for Special Effects.

Ehrenstein, W. H., & Ehrenstein, A. (1999). Psychophysical Methods. *Modern Techniques in Neuroscience Research*, 1325. https://doi.org/10.1007/978-3-642-58552-4_43

Erkelens, C. J., & Collewijn, H. (1985). Motion perception during dichoptic viewing of moving random-dot stereograms. *Vision Research*, *25*(4), 583–588. https://doi.org/10.1016/0042-6989(85)90164-6

Feldmann, I., Schreer, O., Kauff, P., & Nachrichtentechnik, F.-I. (2003). Nonlinear depth scaling for immersive video applications i. feldmann, o.schreer. *Disp*, 1–6.

Garding, J., Porrill, J., Frisby, J. P., & Mayhew, J. E. W. (1996). Uncalibrated Relief Reconstruction and Model Alignment from Binocular Disparities. *Computer Vision — ECCV '96*, 425–438. https://doi.org/10.1007/BFb0015555

Gibson, J. J. (1950). *The Perception of the Visual World*. Boston: Houghton-Mifflin.

Girshick, A. R., & Banks, M. S. (2009). Probabilistic combination of slant information: weighted averaging and robustness as optimal percepts. *Journal of Vision*, *9*(9), 8.1-20. https://doi.org/10.1167/9.9.8

González, E. G., Allison, R. S., Ono, H., & Vinnikov, M. (2010). Cue conflict between disparity change and looming in the perception of motion in depth. *Vision Research*, *50*(2), 136–143. https://doi.org/10.1016/j.visres.2009.11.005

Gottsdanker, R., Frick, J. W., & Lockard, R. B. (1961). Identifying the acceleration of visual targets. *British Journal of Psychology*, *52:1*, 31.

Hershenson, M. (1999). *Visual Space Perception: a primer*. Boston: Massachusetts Institute of Technology.

Hildreth, E. C., & Royden, C. S. (2011). Integrating multiple cues to depth order at object boundaries. *Attention, Perception, & Psychophysics*, *73*(7), 2218–2235. https://doi.org/10.3758/s13414-011-0172-0

Howard, I., Fujii, Y., & Allison, R. (2014). Interactions between cues to visual motion in depth. *Journal of Vision*, *14*(2014), 1–16. https://doi.org/10.1167/14.2.14.doi

Howard, I. P. (2008). Vergence modulation as a cue to movement in depth. *Spatial Vision*, *21*(6), 581–592.

Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth*. Thornhill, Ont: I. Porteous.

Ijsselsteijn, W. a., De Ridder, H., & Vliegen, J. (2000). Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration. *IEEE Transactions on Circuits and Systems for Video Technology*, *10*(2), 225–233. https://doi.org/10.1109/76.825722

Ijsselsteijn, W. a., Seuntiens, P. J. H., & Meesters, L. M. J. (2005). Human Factors of 3D Displays. In T. Schreer, Oliver; Kauff, Peter; Sikora (Ed.), *3D Videocommunication: Alogrithms, concepts, and real-time systems in human-centered communication*. John Wiley & Sons.

Jones, G. R., Lee, D., Holliman, N. S., & Ezra, D. (2001). Controlling perceived depth in stereoscopic images. *Proc. SPIE*, *44*(August), 42–53. https://doi.org/10.1117/12.430855

Kellnhofer, P., Ritschel, T., Myszkowski, K., & Seidel, H. P. (2013). Optimizing disparity for motion in depth. *Computer Graphics Forum*, *32*(4), 143–152. https://doi.org/10.1111/cgf.12160

Kersten, D., Mamassian, P., & Knill, D. (1994). Moving cast shadows and the perception of relative depth.

Max-Planck-Institut Für Biologische …, (June). Retrieved from http://www.kyb.mpg.de/fileadmin/user_upload/files/publications/pdfs/pdf1461.pdf

Laldin, S. R., Wilcox, L. M., Hylton, C., & Allison, R. S. (2012). Motion in depth constancy in stereoscopic displays. In *Proceedings of SPIE - The International Society for Optical Engineering* (Vol. 8288). https://doi.org/10.1117/12.910577

Lang, M., Hornung, A., Wang, O., Poulakos, S., Smolic, A., & Gross, M. (2010). Nonlinear disparity mapping for stereoscopic 3D. *ACM Transactions on Graphics*, *29*(4), 1. https://doi.org/10.1145/1833351.1778812

Leibowitz, H. W., & Harvey, L. O. (1967). Size matching as a function of instructions in a naturalistic environment. *J EXP PSYCHOL HUMAN*, *3*(74), 378–382.

Lipton, L. (1982). Foundations of the Stereoscopic Cinema. *Foundations*, *259*, 258. Retrieved from http://www.getcited.org/pub/102141643

López-moliner, J., Maiche, A., & Estaún, S. (2003). Perception of acceleration in motion-in-depth with only monocular and both monocular and binocular information. *Perception*, *24*, 93–108.

Masaoka, K., Hanazato, A., Emoto, M., Yamanoue, H., Nojiri, Y., & Okano, F. (2006). Spatial distortion prediction system for stereoscopic images. *Journal of Electronic Imaging*, *15*(1), 13002. https://doi.org/10.1117/1.2181178

McKee, S. P., & Welch, L. (1989). Is there a constancy for velocity? *Vision Research*, *29*(5), 553–561. https://doi.org/10.1016/0042-6989(89)90042-4

Mendiburu, B. (2012). *3D movie making: stereoscopic digital cinema from script to screen*. Retrieved from http://books.google.com/books?hl=en&lr=&id=36B5AgAAQBAJ&oi=fnd&pg=PP1&dq=3D+Movie+Making+Stereoscopic+Digital+Cinema+from+Script+to+Screen&ots=r-9P0niLDP&sig=veNcA8hfaAAF-GRYMw5VEuoFkE0

Mueller, A. S., & Timney, B. (2014). Effects of radial direction and eccentricity on acceleration perception. *Perception*, *43*(8), 805–810. https://doi.org/10.1068/p7776

Ni, R., Braunstein, M. L., & Andersen, G. J. (2004). Perception of scene layout from optical contact, shadows, and motion. *Perception*, *33*(11), 1305–1318. https://doi.org/10.1068/p5288

Ni, R., Braunstein, M. L., & Andersen, G. J. (2007). Scene layout from ground contact, occlusion, and motion parallax. *Visual Cognition*, *15*(769429659), 48–68. https://doi.org/10.1080/13506280600646657

Niu, Y., Feng, W.-C., & Liu, F. (2012). Enabling Warping on Stereoscopic Images. *ACM Trans. Graph.*, *31*(6), 183:1--183:7. https://doi.org/10.1145/2366145.2366202

Philbeck, J. W., & Loomis, J. M. (1997). Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology. Human Perception and Performance*, *23*(1), 72–85. https://doi.org/10.1037/0096-1523.23.1.72

Regan, D., & Beverley, K. I. (1979). Binocular and monocular stimuli for motion in depth: changing-disparity and changing-size feed the same motion-in-depth stage. *Vision Research*, *19*(12), 1331–1342. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/532101

Regan, D., Erkelens, C. J., & Collewijn, H. (1986). Necessary conditions for the perception of motion in depth. *Investigative Ophthalmology and Visual Science*, *27*(4), 584–597.

Ritter, M. (1977). Effect of disparity and viewing distance on perceived depth. *Perception & Psychophysics*, *22*(4), 400–407. https://doi.org/10.3758/BF03199707

Rushton, S. K., & Duke, P. A. (2009). Observers cannot accurately estimate the speed of an approaching object in flight. *Vision Research*, *49*(15), 1919–1928.

Sedgwick, H. A. (1986). Space Perception. In J. P. Boff, K. R., Kaufman, L. Thomas (Ed.), *Handbook of human perception and performance*. New York.

smpte.org. (n.d.). SMPTE specs.

Snowden, R. J., & Braddick, O. J. (1991). The temporal integration and resolution of velocity signals. *Vision Research*, *31*(5), 907–914. https://doi.org/10.1016/0042-6989(91)90156-Y

Spottiswoode, R., & Spottiswoode, N. (1953). *The Theory of Stereoscopic Transmission and its Application to the Motion Picture*. https://doi.org/10.1080/01422419908228843

Thx.com. (n.d.). HDTV Set Up.

Trewhella, J., Edwards, M., & Ibbotson, M. R. (2003). Sensitivity to the acceleration of looming stimuli. *Clinical Experimental Ophthalmology*, *31*(3), 258–261.

Trivedi, H. P., & Lloyd, S. a. (1985). The role of disparity gradient in stereo vision. *Perception*, *14*, 685–690. https://doi.org/10.1068/p140685

Wallach, H., & Zuckerman, C. (1963). The Constancy of Stereoscopic Depth. *The American Journal of Psychology*, *76*(3), 404–412.

Wang, C., & Sawchuk, A. A. (2008). Disparity Manipulation For Stereo Images and Video, *6803*, 1–12.

Wattie, J. (2012). Stereoscopic Roundness depends on Four Parameters.

Werkhoven, P., Snippe, H. P., & Toet, a. (1992). Visual processing of optic acceleration. *Vision Research*, *32*(12), 2313–2329. https://doi.org/10.1016/0042-6989(92)90095-Z

Woods, A., Docherty, T., & Koch, R. (1993). Image Distortions in Stereoscopic Video Systems. *SPIE's Symposium on Electronic Imaging: Science and Technology*, *36*(February 1993), 36--48. https://doi.org/10.1117/12.157041

Yamanoue, H., Okui, M., & Okano, F. (2006). Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images. *IEEE Transactions on Circuits and Systems for Video Technology*, *16*(6), 744–752. https://doi.org/10.1109/TCSVT.2006.875213

Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research*, *33*(18), 2685–96.