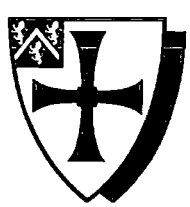
An assessment model and implementation of stereo image quality

Raincock, Samantha

An Assessment Model and Implementation of Stereo Image Quality

by

Samantha Raincock



Submitted in conformity with the requirements
for the degree of Master of Science
Department of Computer Science
University of Durham

Dedication

I give thanks to the following:

- To my Family:
  - To my all my family and pets.
  - In memory of Granddad Atkinson.
  - In celebration of Granddad Raincock in his awesome fight against cancer.
- To members of Staff at Computer Science, Durham University:
  - Liz Burd - Without whom I doubt this thesis would exist.
  - Brendan Hodgson - For continued support and a unique sense of humour.
  - Sarah Drummond - For continuous support throughout my time at university.
  - Edwin Brady - For compiling a Latex thesis format appropriate to the university requirements.
- To my sponsors:
  - Sharp Research Laboratories Europe
- To my friends:
  - Powlo, Ian McKinley, Jonathan Gough and David Johnson who have advised me through the many phases of my thesis journey.
  - To all my friends - your guidance and support has made this possible.

Declaration

The material contained within this thesis has not previously been submitted for a degree at the University of Durham or any other university. The research reported within this thesis has been conducted by the author unless indicated otherwise.

Copyright Notice

Contents

List of Figures

List of Tables

Abstract

An Assessment Model and Implementation of Stereo Image Quality

Samantha Raincock

In the past decade, many display hardware manufacturers have initiated research into the construction of stereo display devices. Currently, the use of such displays is limited to the computer-aided design; research, military and medical applications. However, it is anticipated that as display hardware becomes cheaper, gaming companies and desktop application software developers will realise the potential of using stereo to provide more realistic user experiences.

To provide realistic stereo user experience it is necessary to utilise good quality stereo images in addition to suitable hardware. The growth of the Internet has resulted in an increase in the availability of stereo images. However, most have been captured using uncontrolled procedures and have questionable quality. The quality of stereo images is important since the viewing of poor quality stereo images can result in adverse viewing effects.

A formal definition of stereo quality has not been achieved in current day research. This means that the factors which cause a stereo image to be perceived as poor quality have not been defined nor is a system available to detect its occurrence. This thesis attempts to address this problem by postulating a definition of stereo image quality based on detecting level of excess disparity levels, intensity differences and the occurrence of frame cancellation. An implementation system able to detect these identified factors is discussed and formulated.

The developed system is utilised to test 14 stereo images of varying quality levels. The results of these tests are reported and are used to evaluated and refine the system. Using this image analysis, benchmarks for natural intensity difference in images, changes due to JPEG compression and comparisons with generated and groundtruth disparity maps are formulated. Additionally, an

analysis of the system is performed, to determine the model and implementation weaknesses including areas of further work.

Chapter 1

Introduction

1.1 Overview

The use of stereo images in the medical profession, computer aided design (CAD), military intelligence and gaming has grown in recent years. This is due to a growing knowledge of the benefits of observing stereo images. In some cases, the use of stereo has revolutionised the ability to provide an accurate assessment of an observed scene. One example of this is the diagnosis of glaucoma in a diabetic eye. The use of stereo by specialist ophthalmologists has allowed the earlier diagnosis of glaucoma which may be used to save the eye sight of hundreds of patients by providing more immediate treatments.

Over the past 5 years there has been an increasing interest by display manufacturers in the production of stereo display systems. This has primarily centred on the production of autostereoscopic display systems. These displays allow users to perceive disparity without the need for any additional specialist equipment beyond that of the display. This makes the display much less intrusive to the user. Autostereoscopic displays have been incorporated into LCD display panels, laptops, wide screen display devices and mobile telephones and are currently readily available.

The introduction of autostereoscopic display systems and the growing hard-

ware manufacturers' interest in stereo displays has seen an increase in the production of stereo images. Stereo images can be downloaded from many sites on the Internet usually captured by amateur stereo photographers. There are also professional and controlled stereo images captures available.

The definition of quality in reference to stereo images is a new research area. There is no quantitative or qualitative definition of what constitutes a good quality stereo image. This means that it is currently not possible to assess if a stereo image is of a high enough quality for human viewing or even what such an image would require.

The observation of stereo images containing defects may lead to adverse side effects and/or the inability to fuse the stereo image as a single perception. Such stereo images should be avoided. However, currently there are no stereo image assessment systems which can examine a stereo image and analyse its suitability for human viewing. Such a system would be of use to users, display manufacturers and stereo image generators alike.

This thesis attempts to bridge this gap in current knowledge. It aims to provide a basic definition of quality in stereo images and to produce an assessment system based on the formulated definition. The system will be able to detect certain attributes in stereo images and produce an analysis of the findings. Such a system could be used to assess stereo images for their viewing quality and prevent the observation of stereo images which may cause adverse side effects.

1.2 Thesis Aims

The follow is a list of aims of the thesis which will ultimately be used as success criteria:

- To provide an overview of human vision including the history and current

research into stereo human factor analysis.

- To provide an introduction to stereo display technologies and image capture.
- To provide a geometric model of stereo perception on autostereoscopic display systems.
- To develop a basic model for assessing the quality of stereo images.
- To develop an implementation of the stereo quality model able to analyse a subset of stereo images.
- To evaluate the stereo quality assessment system using a set of test images.
- The evaluation of the model and implementation including possible refinements, pattern recognition and a critical analysis.

1.3 Thesis Organisation

The remainder of the thesis is organised as follows:

Chapter 2 reviews the general background information concerning human vision, human stereo vision, depth cues, stereo display hardware and problems encountered by the use of stereoscopic displays.

Chapter 3 initially performs an analysis of the current stereo image generation techniques and available assessment systems. The chapter then goes on to examine the problems which may be encountered when viewing the world in stereo specifically when using artificial display systems and images. The chapter then provide an in depth analysis of a subset of the associated problems, attempting to overview current human factors research. This will include measured limits and observations.

Chapter 4 discusses the proposed model for assessing of stereo image quality. An in depth discussion will be made regarding the possible model processes and the decisions made regarding the finalised form. Each of the included processes will be concisely summarised and explanations of the design choices made.

Chapter 5 discusses an implementation plan of the model discussed in Chapter 4 for building a system able to assess quality in stereo images. The chapter will examine the technical difficulties and limitations in the production of an implementation and detail the anticipated inputs and outputs of the system.

Chapter 6 analyses the results obtained from testing the developed system using a selection of stereo images. The chapter will then analyse the results for each image and image sets and present the findings from the analysed images.

Chapter 7 will draw any conclusions found from the analysis of the results detailed in Chapter 6. It will then discuss any limitations, problems and areas of improvement which may have effected the results obtained. To conclude the thesis, a summary of possible further work will be made.

Chapter 2

An Introduction to Human Vision and Stereo Viewing

This chapter will give an overview of the literature regarding human vision and human perception. It will examine how humans perceive the world including both monocular and binocular vision. Specifically, the chapter will examine the depth cues utilised by the human visual system and how these are used to create a sense of depth. It will provide a summary of their occurrence and significance in scene perception. The chapter will conclude by examining stereo display systems including an introduction to their use and the geometrical modelling of their stereo image projection.

2.1 Human Vision

The human visual system is an exceedingly complicated structure that starts with the eyes as a light detector. The eyes then transmit the detected light as a signal which is interpreted by the brain. The brain uses this information to bring about a perception of its observed environment. The processes involved in bring about this perception are not fully understood, however, theories exist regarding the structure of the brain and the resulting perceptions. This section will concisely summarise the main elements of the system and the resulting perceptions.

2.1.1 An Introduction to Human Vision

When an observer views the world around them, they can do so with either a single eye or both eyes simulatneously. Viewing the world with only one eye is referred to as *monocular vision* and viewing the world with both eyes is referred to as *binocularly vision.*

The eye to eye distance is termed the *interocular distance.* When humans observe a scene, the interocular distance causes a level of horizontal displacement to be present between the two images observed by each eye. This can be demonstrated if the reader selects a point in space and closes one eye and then switches to the other. The observer will see a horizontal shift in the position of the observed object. The displacement observed is known as *horizontal disparity.* Often within the literature the term disparity is referred to as *parallax*, *stereo* and *binocular disparity* (due to the disparity cue is formed by observing the world using *binocular vision*). These terms can be used interchangeably. *Stereopsis* is the ability of the human visual system to perceive depth in an observed stereo pair due to the disparity between the two images.

The process by which the images from the eyes are brought together as one in the brain, is known as *fusion* [56]. Fusion can only occur when the two perceived images are similar. When the images differ the result may be an

inability to perceive the scene as a single fused perception and the possibility of adverse side-effects. This can be demonstrated if the reader places their eyes at a door frame so that one eye perceives one room and the other another.

The point of fixation [50] in binocular viewing is a single point within a visual scene where both eyes are *focused* and *converged* at any given time. Focusing on an object involves the eye observing the object as a sharp and undistorted image. Focusing happens with little conscious interaction by the observer. The eyes refocus just like a camera lens to achieve a sharp image at the point of fixation. This process is often referred to as *accommodation.* Movement of the eyes occurs because unlike a camera lens the human eye is able to rotate and is not in a fixed position. The rotation of the eyes to observe something within a scene is known as *convergence.*

2.1.2 Physiological Structure of the Human Eye

The eye [32][69][50][29] is the external organ responsible for capturing light. It enables an observer to perceive their surroundings. The eye is a structure containing three externally visible parts: the *sclera*, *iris* and *pupil.* The sclera is the white external area of the eye. It is not directly involved in vision but performs a support and structural maintaining role. The iris is a circle of muscle that controls the size of the pupil; it is the visible coloured part of the eye. The pupil is the central aperture of the iris. It is through the pupil that light passes into the eye and the processes of perception begins. Figure 2.1 demonstrates the simple structure of the human eye.

The surface of the external eye is covered in a transparent layer known as the cornea. The cornea is the most powerful focusing mechanism in the eye; it has more *refracting power* (the ability to bend light) than the eye's lens. Approximately 80% of the total refraction of the light entering the eye occurs at the cornea. In combination, the lens of the eye and the cornea allow the eye to focus on objects in a perceived scene.

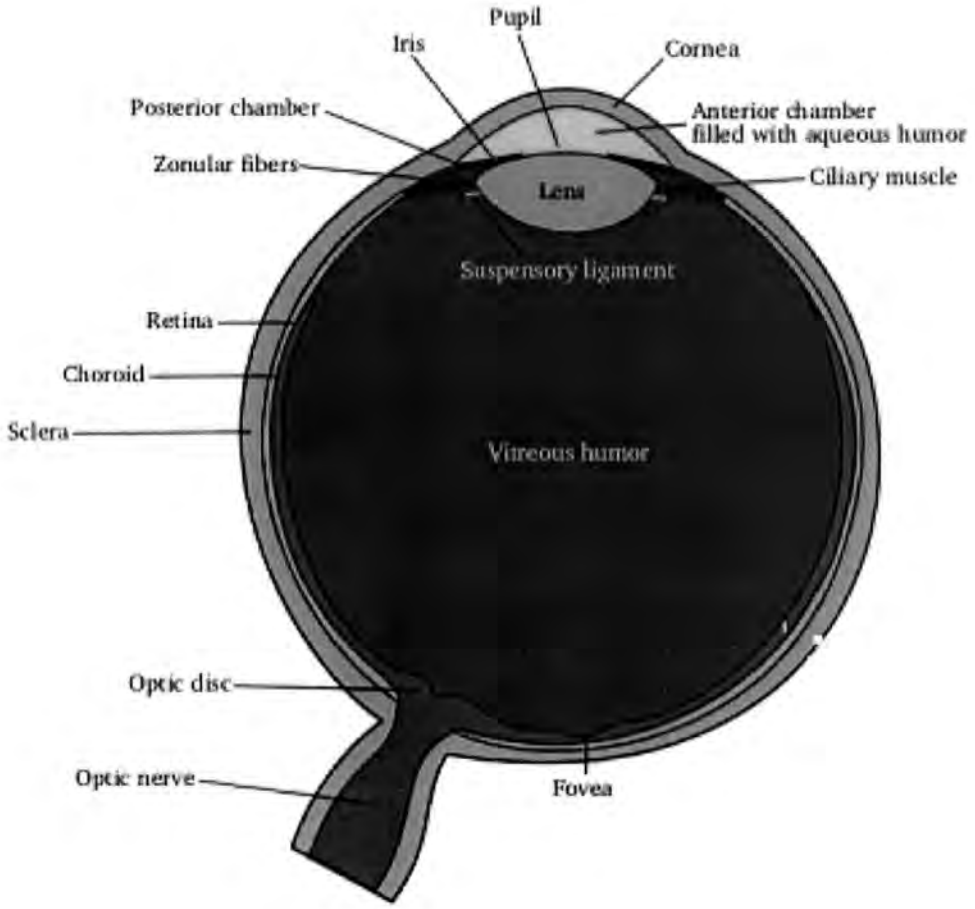


Figure 2.1: The Human Eye [69]

When light passes through the cornea, it enters the aqueous humor. The aqueous humor is positioned between the cornea and iris, in the anterior chamber and is a jelly like substance. Light then enters the posterior chamber, which is positioned between the iris and the lens, where it passes through the pupil to encounter the eye's lens. The eye's lens is held in place by ligaments and muscles. The contractions of these change the shape of the lens allowing accommodation to occur. The light then passes through the lens and continues through the vitreous humor and finally collides with the retina. It is the structure of the retina which facilitates the translation of the light detected by the

eye into a perception.

The eye's retina is a complex structure. It is a light sensitive thin layer covering approximately 65% [69] of the internal structure of the eye. It contains *photosensitive cells* [29](rods and cones) that are able to detect and react to light colliding with them. In the middle of the retina is a structure called the *fovea.* The fovea is the point on which a functional eye focuses light. It is at this location that the finest *resolution* (ability to detect detail within an image) can be achieved. Additionally, the retina contains a region known as the *optical disk.* It is through this structure that the optical nerves transmit the signals from the eyes to the brain for interpretation. In this area, there are no photosensitive cells and hence any colliding light will remain undetected. Often the brain 'fills in' the perception of a scene in this area so that in normal viewing conditions it remains unnoticed. This area when observed under specialised viewing conditions or experiments is known as the *blind spot.*

An average human observer has an interocular distance in the range 63-65mm [61][41] but natural variation can demonstrate a range between 63-70mm. Observer's interocular distance rarely exceeds 85mm [95]. The variations found in the interocular distance are discussed in detail by Dodgson [20].

The interocular distance is usually assumed to be horizontal displacement only, however, in some human observers the eyes can be misplaced and a vertical difference can also be present. This thesis uses the term interocular distance to refer to the horizontal displacement only and will take a value of 65mm to be the average interocular distance for an observer.

Discussions on the anatomy and physiology of the eye are comprehensively provided in [56][29][69][18][32][101][50].

Common Visual Defects

Defects in human vision are common, affecting a large proportion of the world's population. WHO [74] estimated that more than 161 million people were visually impaired in 2002 and over half of all people have some form of problem with their vision. Approximately 12% are stated as having some form of dysfunction in their binocular vision [14][31].

Defects in Monocular Vision

Discussed below are common defects found in the vision of the eyes. The deficiencies can occur in one eye or in both and can be different in each. They can cause problems with the observers ability to focus scenes both monocularly and/or binocularly [29][72]:

- **Myopia**
  The light entering the eye is bent too much or the eyeball is too long. This results in an observer being unable to focus effectively and hence experiencing blurred vision especially for distant objects.

- **Hyperopia**
  The light entering the eye is not bent enough or the eyeball is too short. This results in an observer being unable to focus effectively and hence experiencing blurred vision especially for objects viewed close to the eye.

- **Astigmatism**
  Astigmatism usually occurs when the eye's cornea has an irregular shape; being more skewed in shape in one area than another. This results in an observer experiencing blurred vision for all viewing distances.

In addition, eye and brain disorders may also affect the ability of humans to view the world binocularly. This results in an inability to perceive the world with both eyes or to fuse the two images together (stereopsis) to form an overall single perception of the observed environment. Various diseases and conditions can cause binocular vision impairment, giving those afflicted only a limited degree of binocular vision function or none at all. Binocular vision impairment

can cause a variety of problems that can present themselves individually or in combination. These ultimately degrade the ability of an observer to perceive 3 dimensional (3D) depth in observed scenes. Factors causing binocular vision impairment include the inability to eye track objects; to fuse objects; to perform stereopsis and to converge onto a single fixation point.

There are various conditions and circumstances that can cause binocular visual impairment including:

- **Diopia**
  Diopia (double vision) causes two separate images to be perceived simultaneously when viewing a scene binocularly. It often results from a failure to converge the eyes correctly onto a single point of fixation.

- **Strabismus**
  Strabismus [72] is more commonly known as crossed-eyes. In such a condition, a person is unable to converge their eyes correctly so that they both focus on the same point in visual space. This condition can give symptoms of diopia or *suppression* (the brain discarding the image of one of the eyes).

- **Amblyopia**
  Amblyopia [72] is more commonly referred to as lazy-eye. The condition occurs when normal vision is not present during childhood. Usually one eye develops good vision and the other does not, leaving the malsighted eye as being amblyopic. If the condition is corrected in childhood, it is possible for normal binocular vision to form but if it is left untreated, it can result in permanent binocular vision loss.

Diseases and vision problems associated with the eye and possible treatments are discussed in [29][17][43] and a comprehensive dictionary of eye diseases is given in [102].

The development of binocular vision in the brain primarily occurs before the age of 8 years. Prior to this age the brain structures grow and eventually

develop to allow for adult vision. Correction of problems in vision (single eye conditions and binocularly fusion problems) prior to this age is essential to ensure that children grow up with the ability to perceive the world binocularly. Similarly, binocular experiences which are detrimental to childhood binocular visual development should be avoided since it is possible that such experience could additionally cause permanent binocular visual problems as an adult.

Children's binocular vision is also different to that of an adult due to the difference in the interocular distance size. This means that the horizontal displacement between the two observed images is typically smaller in children than adult observers. This results in children perceiving the world with a greater level of depth than adults.

2.1.3 An Overview of Physiology of the Human Visual Brain

Discussed in [56][29][50][75] are the processes involved in physiologically utilising the signal initiated at the retina to bring about a scene perception. The structure of the human brain and the visual processes causing the perception involve highly complicated processes. These are described in varying degrees of detail within [56][29][50][75][26][37]. A detailed discussion of the physiological theories which describe how human perceptions and stereopsis occurs is beyond the scope of this thesis. Hence only an overview will be given for human stereopsis.

The Physiology of Binocular Vision

The processes involved in converting the disparity information encoded at the retinas to form an overall perception within the brain is known as stereopsis [56][29][30]. Stereopsis is the process by which the brain formulates a meaningful correspondence of the two images. There is no full explanation of how stereopsis occurs in the brain, however, theories and attempts have been made

to try and explain it from experimentation results and observations.

One theory involves the analysis of how corresponding points are fused in the brain. Corresponding points are points within a scene of an object which is located at different physical locations on the retina in each eye. This difference is due to the presence of the interocular distance.

When the eyes fixate on a point all areas at the same depth from the observer form a theoretical circle known as the *horopter* [41] (often referred to as the Vieth-Muller circle [56]). This is demonstrated in Figure 2.2.

A theory of correspondence in the human visual system states that all the points located on the horopter, for any given fixation point in a scene, are located on corresponding points on the retina of each eye. Those points are said to be retinally fixed. All other points not located on the horopter are located on non-corresponding points. The theory postulates that corresponding and non-corresponding points are used by the brain to bring about image fusion and hence a stereo perception. Correspondence can be thought of as a matching process between the two images, received by both eyes, within the brain. The fusion of the two images gives the visual system a sense of depth. This is achieved by the degrees of disparity between non-corresponding points and the point of fixation [30].

2.2 Perceiving Depth

Depth cues are indicators that provide information interpreted by the visual system to give an observer a sense of depth in an observed scene. These depth cues can be obtained by both monocular and binocular viewing. Figure 2.3 [85] demonstrates a basic summary of the depth cues used by the visual system to perceive depth.

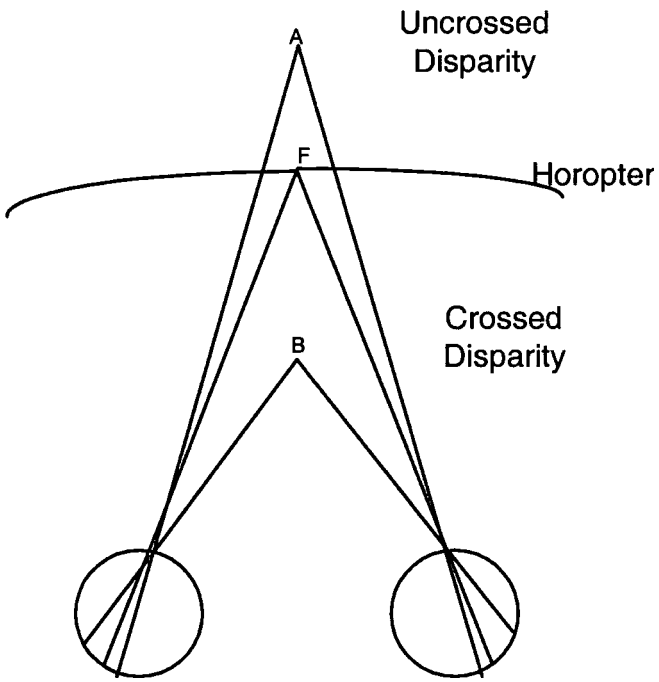


Figure 2.2: Horopter

2.2.1 Depth Cues

Binocular disparity is not the only cue present to allow an observer to gain a sense of depth in a scene. There may be many other cues that aid an observer in establishing depth. It is the combination of these cues that brings about an overall perception of depth, size, environment etc.

It is stated in [56][29][16][25] that the visual system uses four different categories of visual cues to derive a perception of depth from a viewed scene. These include the following:

- **Oculomotor Cues (Physiological Cues)**
  The ability of the visual system to sense the position of its eyes and the tension in the eyes muscles to gain a sense of where the eyes are focusing and accommodating. This depth cue can be used by the visual system in both monocular and binocular viewing.

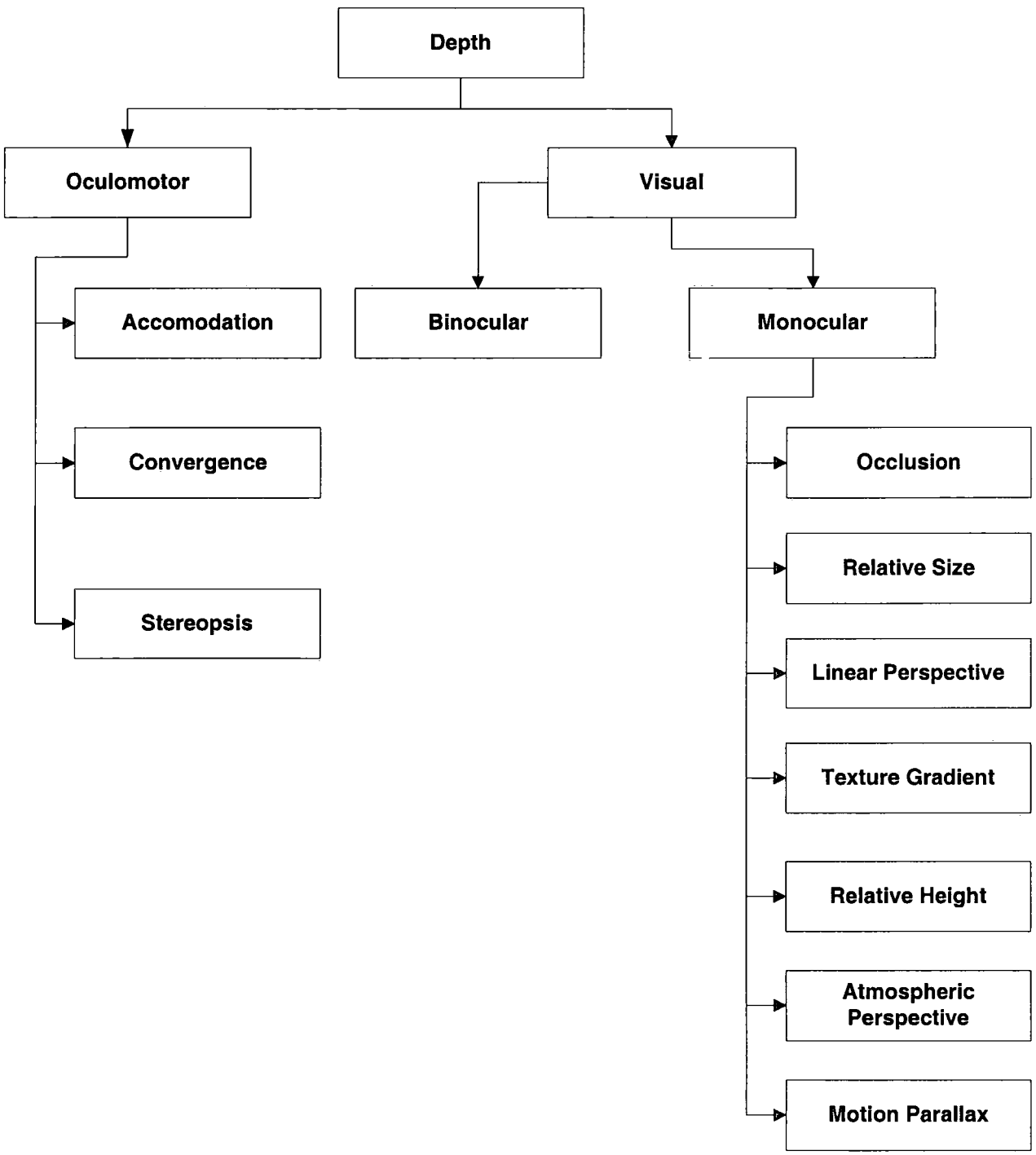


Figure 2.3: Human Depth Cues

- **Pictorial Cues**
  The depth cues obtained from viewing a scene monocularly. These cues are present in most observed scenes.

- **Movement Cues (Kinetic Cues)**
  The depth cues created by either the movement of the observer in relation to an observed scene or the movement of objects within the scene in relation to an observer. The cue is achieved from a change in viewpoint.

- **Binocular Disparity**
  The depth cue provided due to the left and right eyes receiving slightly different viewpoints of a scene. This is due to their slightly different positions in visual space. This cue is only present when an observer views the world with both eyes and with functional binocular vision.

Oculomotor Cues

Oculomotor cues [29][16] include convergence and accommodation:

- **Convergence**
  The rotation of the eyes to facilitate the fixation of the eyes onto an object within a given scene.

- **Accommodation**
  The change in the focal length of the lens of the eye to allow it to focus on objects at varying depths.

The brain controls the process of convergence. The oculomotor muscles are initially used to reorient the eyes to bring about fusion roughly on the desired

fixation point. This is then fine-tuned to an orientation with only a very small degree of error. This error is known as the *fixation error*. This error can cause an observer to fixate on a point either slightly in front or behind the fixation point and has been found experimentally to depend on the distance to the point of fixation. The results [46] suggest that closer points cause a larger degree of fixation error than those at a further distance from the observer. A small degree of fixation error is tolerated within the visual system and fusion is still possible in its presence.

Oculomotor cues are present in both monocular and binocular viewing. In monocular viewing the necessity of fixation onto an object is still required. The reader will experience this when they close one eye and move the observing eye. The closed eye will move in sequence to the observing eye and effectively both eyes are still converging. The convergence and accommodation of the eyes on to a point of fixation is stated to provide a disparity angle on the retina that is interpreted by the visual system to represent a specific disparity within a given scene.

Pictorial/Empirical Cues

Pictorial cues [56][29][41][60][16] are provided with or without the presence of binocular vision. They are provided purely monocularly and are present in the majority of observed scene. A selection of pictorial cues utilised by the visual system are the following:

- **Perspective**
  Objects that are further away appear to be smaller than those that are closer.

- **Relative Size**
  The observation of a smaller image on the retina than another retinally perceived object results in the visual system deducing that the smaller image is located at a larger depth than that of the larger object.

- **Relative Height**

Objects with their base located higher in the visual field are perceived as more distant than those located lower in the visual field. This is because the visual system is accustomed to observing distant scenes in the horizon.

- **Size of Known Objects**
  Knowledge acquired by previous viewing experiences gives observers a knowledge base of already encountered objects sizes. When unknown (or known) objects are observed against already encountered objects an approximation of their relative size is formulated by the visual system.

- **Occlusion/Interposition**
  An object, which blocks another is assumed to be in the foreground with respect to the blocked object. This gives a suggested depth ordering within a scene. The blocked object is said to be occluded.

- **Aerial/Atmospheric Perspective**
  The atmosphere in which light travels effects the perception of distance. Weather conditions such as fog, rain and dust storms can greatly affect depth perception. As the light travels through the medium of air, when it encounters particles they cause the light to scatter. This causes the light to lose its saturation meaning that sharp edges of objects become blurred and the wavelength of the light is shifted towards the visible blue spectrum. This effect is also observed with changes in altitude and humidity. This results in a change in the perceived distance to objects in a scene since the visual system becomes adapted to its usual environment. This is specifically apparent when an observer moves from one set of conditions to another and can result in observers initially misinterpreting depth.

- **Reflections of Light**
  The way light behaves when it encounters particular surfaces can be utilised by the visual system to determine the surfaces curvature and positioning.

- **Shadows**
  Shadows may aid the occlusion cue by the act of them casting onto other objects. This may provide an additional depth ordering cue.

- **Relative Brightness**
  Closer objects are usually perceived as brighter than those that are more distant. This is because the light reflecting from distant objects will encounter more scattering or change than closer objects.

- **Texture Gradient/Detail**
  Elements that are equally spaced within a scene appear to be packed closer together as the distance from an observer is increased. This is a simple application of the perspective cue, however, its result is that equally spaced structures are interpreted as containing depth.

A summary of some of the pictoral cues present in flat images is demonstrated in Figure 2.4.

Movement Produced Cues

The effects of motion on depth perception [29][16] involve two types of cue:

- **Motion Parallax**
  Two objects moving at the same velocity but at different distances from an observer will be perceived to move at different velocities. Distant objects are perceived to move slower than those located closer to the observer. By integrating temporally separated view points, over time an observer can get a simulation of depth just like that created from the difference in the view point caused by the interocular distance in binocular vision.

- **Deletion and Accretion**
  If an observer moves horizontally in respect to an observed scene this

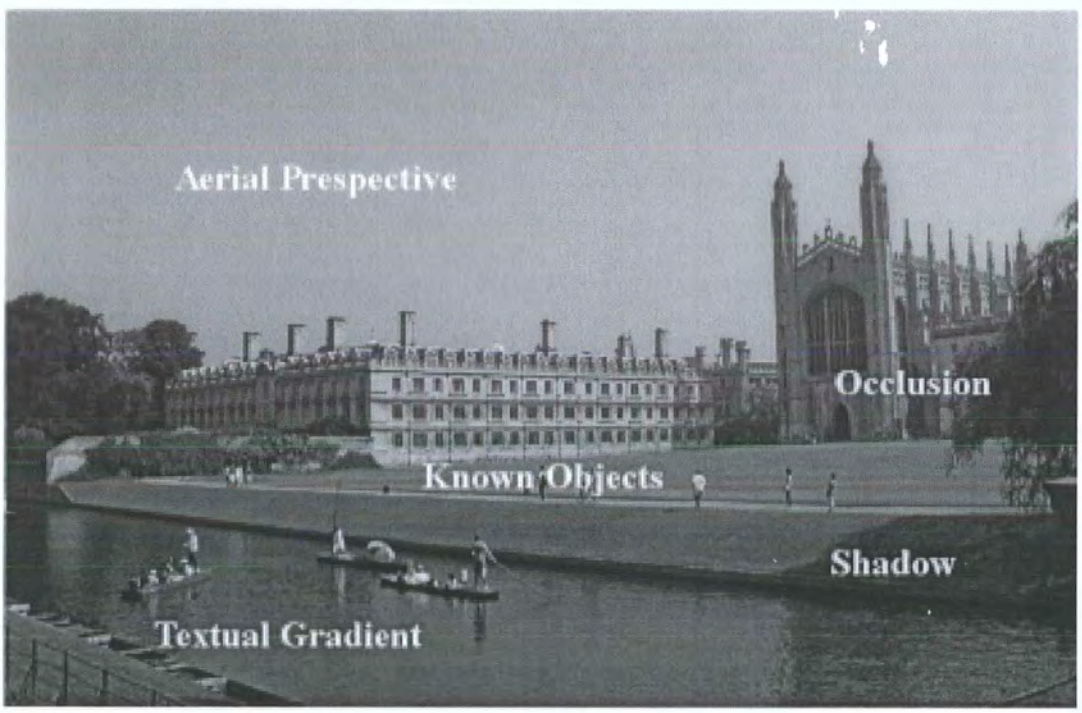


Figure 2.4: 2 Dimensional Depth Cues - Picture of Cambridge Kings College, Photographer - Sam Raincock

results in objects at varying distances appearing to move relative to one another. When the observer moves in one direction objects closer to the observer delete (temporally occlude) and then accreted (become re-exposed) when the observer moves in the opposite direction. A similar situation arises when objects are moving and the observer remains stationary.

When an observer or objects in a scene move, the image projected on the retina also moves. However, there are many cases where an object moves on the retina but the resulting perception is stationary or when objects remain stationary on the retina but the perception formed is that of motion. These phenomena are known as *stroboscopic movement.* Additionally, when we move in an environment the scene's images move across the retina. However, we may not perceive this as a scene in motion. The brain rationalises the movement as its own person and not the objects within the observed scene.

Binocular Cues

Binocular vision is defined by Holliman [41] to be the ability of an observer to use two eyes to perceive the depth cues available within a 3 dimensional world. Binocular disparity [29][43][50][82] is brought about by the differences in viewpoint of the left and right eyes due to the interocular distance. The perception of depth attributed to binocular disparity is due to the difference in position of non-corresponding points to corresponding points located on the retina. This is utilised by the brain to give a sense of depth in a process known as stereopsis.

Vertical Disparity

Discussed within Matthews [64] vertical disparity is an additional valid depth cue to horizontal disparity. Vertical disparity involves a displacement in corresponding regions of the left and right image as a vertical shift. It can be present in addition to horizontal disparity or as the single binocular cue. Theories have been proposed regarding the physiological arrangement of vertical

disparity in the brain as a weaker form of horizontal disparity. The major evidence for the existence of vertical disparity effecting human depth perception is The Wallpaper Illusion [63]. However, excess vertical disparity is known to cause visual disturbances and should be avoided.

2.2.2 The Influence of Depth Cues

The depth cues discussed in the preceding section describe how each individual cue is utilised by the human visual system to obtain a sense of depth. However, the sense of depth achieved is both limited by the cue's active ranges and their interactions. The cue ranges are limited to the following parameters [29][50]:

- **Height in the visual field** - Maximally effective at 2m and decreases in effectiveness up to 100m.

- **Aerial Perspective** - Effectiveness increases from a distance of about 100m - 2Km and then decreases at distances greater than several kilometres.

- **Motion Perspective** - The effectiveness is dependant on the velocity of the objects, however, in general the effectiveness decreases at distances over 100m.

- **Binocular Disparity** - The cue provides the greatest information for very close objects and decreases in effectiveness at about 10m.

Table 2.2.2 demonstrates the depth ranges of each of the examined depth cues [29].

	Depth Ranges		
Depth Cue	0-2m	2-30m	30+m
Occlusion	•	•	•
Relative Size	•	•	
Accommodation and Convergence	•		
Motion Parallax	•	•	
Binocular Disparity		•	•
Relative Height		•	•
Atmosphere Perspective			•

Table 2.1: Depth Cue Ranges

Limited research has been performed regarding the influence of each of the depth cues in the final perception result, however, Crascic [16] discusses weightings in these categories. He rates them from the largest to the smallest influence for the Oculomotor cue and pictorial cues as follows:

- Oculomotor Cues: Convergence → Accommodation.

- Pictorial Cues: Interpolation → Linear Perspective → Texture Perspective → Aerial Perspective → Relative Brightness → Shadows.

2.2.3 Cue Conflicts

Cue conflicts [106][66] may cause ambiguity and confusion. They exist when one or a number of depth cues imply that the depth is at a particular level and one or more other depth cues contradict this.

Conflicts between the Oculomotor Cues

An interesting cue conflict can occur between the two oculomotor cues; the convergence and accommodation measurements within the eye. This conflict

occurs when the eyes accommodate and converge on to different points within 3D space. In normal functional binocular vision, convergence and accommodation are tightly coupled; observers usually converging and accommodating onto a single point; the fixation point. When the visual system converges and accommodates onto different points, research [106][66] suggests that such a conflict can cause eyestrain and confusion. However, many everyday optical devices can disturb the tight coupling of these cues. For example, the wearing of eyesight corrective spectacles [16][108]. This does not seem to cause most users difficulties. It is also contrary stated [16][11] that the oculomotor cue conflict does not cause eyestrain but can cause perceptual errors. However, in general it is accepted that a conflict in the oculomotor cues should be avoided where possible.

Cegalis [11] performed experimentation using prisms to induce a conflict between the convergence and accommodation cues and determined that initially a change in perception was observed. However, the observers quickly adapted and showed no significant after effects. Valyrus [98] also performed experiments measuring the effects of cue conflicts between convergence and accommodation and determined that the discrepancy between the two should be no more than a 1.6^o difference between the point of convergence and accommodation for visual disturbance to result.

Binocular and Monocular Occlusion Conflicts

It is possible in a limited selection of real world visual encounters for an observer to experience a conflict between the monocular pictorial cues and those obtained binocularly from stereopsis. If a conflict arises between the weaker pictorial cues and binocular disparity then the resulting perception will likely favour the perception established through binocular disparity. However, if there is a conflict between the purported depth from occlusion and binocular disparity then occlusion will always prevail. In this sense, occlusion is said to be a far stronger depth cue than even that of binocular disparity.

Shape Consistency

It is possible that the shape of observed objects remains constant even when viewed from different orientations. This means that the object's shape changes on the retina but the visual system perceives them to remain in a constant orientation. This is known as shape consistency [29][47]. There are two types of perceived shape consistency within the visual system:

- **Perfect Shape Consistency**
  The observer sees the same object shape no matter what angle it is observed, even though at different viewpoints a different image is created on the retina.

- **Partial Shape Consistency**
  The observer perceives the object shape similar to the image on the retina but with slight changes. The perception is skewed towards the perception that would be produced by perfect shape consistency.

Size Consistency

Size consistency [29][50][25] involves perceiving the size of objects on the basis of the size of the projection onto the retina. When viewing a scene containing an object of known size, the size perception of the observed object remains the same even though the size of the image on the retina changes. This can be in contradiction to the visual angle that an object projects onto the retina since this is dependant on the size of the object and of the distance the object from the observer.

Our perception of size is based on a consistency scaling mechanism that supplements the information regarding an object with information about its perceived distance. Size distance scaling operates according to Emmert's law. Hence, size perception relies on other depth cues and object knowledge for its formulation. A conflict in the perceived size of an object on the retina, additional depth cues and object knowledge can cause a change in the overall perception of depth within a scene. It is possible to create visual illusions using such a

visual system 'trick'. An example of this is a common scene of a toy car made to look like a full size car due to the surround objects.

Binocular Conflicts

Binocular conflicts occur when there is a difference between the information provided by the left and right eye. This can result in a status of *suppression*, *superimposition* or *binocular rivalry* [56].

Suppression causes the 'switching off' of one eye's image so that the suppressed eye's image is unseen. The observer effectively perceives the world through a single unsuppressed eye.

Superimposition results in one eye's image being presented over the top of the other. This results in the observer perceiving both images either totally superimposed or superimposed in regions of the image.

Binocular rivalry describes a situation of alternating suppression between one eye's image and the other. Of the three described binocular conflicts binocular rivalry causes the greatest level of perception disturbance. This is due to the images effectively switch from one eye's image to the other either in the whole image or regions of the image.

There are two theories on how fusion is performed within the human visual system. These include *Suppression Theory* and *Fusion Theory*. Suppression theory states that observers only perceive their environment through one eye at a time and the images alternates fast between the two. Fusion theory states that we see through both eyes simultaneously.

A state of binocular rivalry [58] is usually caused when there is a strong conflict of orientation, colour or contrast in binocularly viewed images. The result is that only one image is observed and the other completely invisible. This situation continues to alternate between entire images or image regions [3].

This behaviour supports suppression theory. Wheatstone [112] observed that binocular rivalry only occurs when the two images are not similar, however, fusion is favoured and binocular rivalry only occurs when fusion breaks down. This behaviour supports the fusion theory.

Binocular orientation rivalry is observed when one eye is presented with a vertical line and the other a horizontal line of the same proportions and aligned so that they reside on corresponding points within the image [56]. Binocular contrast rivalry [36] occurs when there is a large difference in contrast between the corresponding left and right images. No such rivalry exists when there is just a small degree of change in the contrast levels between the two points. In such cases the human visual system treats them as similar and still brings about fusion.

Monocular Rivalry

Monocular rivalry [3][5] produces the same alternating results as experienced in binocular rivalry, however, the rivalrous conditions are produced in a single image. Monocular rivalry seldom occurs in natural viewing situations but an example of its occurrence is an image containing a red grating placed orthogonal to a green grating within the same image. In such situations, similarly to binocular rivalry the image alternates between the two perceptions. In this case, the perception will alternate between the red and green gratings. This leads to the interesting question as to if monocular rivalry is a factor in binocular visual comfort and more importantly how and why it occurs. Predominantly these questions remain unanswered.

2.2.4 Perceiving Depth - Monocular Vision versus Binocular Vision

Monocular cues exist solely from the information obtained from each eye singly or from the observation of a scene through one eye. There is no fusion in the

brain of the information from both eyes. Hence, monocular cues are available to individuals with deficiencies in their binocular vision as well as those with functional binocular vision. Binocular vision allows the fusion of the information obtained from both eyes to give a greater sense of depth and scene positioning.

The additional benefits [41] provided by binocular vision are as follows:

- A more accurate depth judgements due to the additional information provided through stereopsis.
- Spatial localisation becomes easier as the difference in viewpoint allows the location of objects to be achieved easier.
- Surface material perception, for example gems, transparent and shiny surfaces, provide the visual system with a number of depth perception problems. Using monocular cues alone can make it difficult for the visual system to resolve the scene depth. The additional cue provided by binocular vision can aid the determination of depth and resolve conflicts that may be present when observing such surfaces.
- Judgements of curved surfaces can be difficult due to the possible presence of warped shadows and reflections. Binocular disparity can aid resolving conflicts and anomalies that may be present and aid in giving the observer a more accurate perception of the scene depth and ordering.
- A larger visual field is provided due to each eye having a slightly different viewpoint of the world. With eye movement the monocular field for each eye is about 150^o and with both eyes, it is at least 180^o wide.
- Binocular summation occurs in the overlapping section of a binocularly viewed image. This area is known as the *binocular visual field.* The presence of two viewpoints of the same scene may give improvements to the whole perception of the scene. The improvements may be very small but it will give slight advantages in the detection of scenes within limited lighting conditions as well as the ability to detect camouflage.

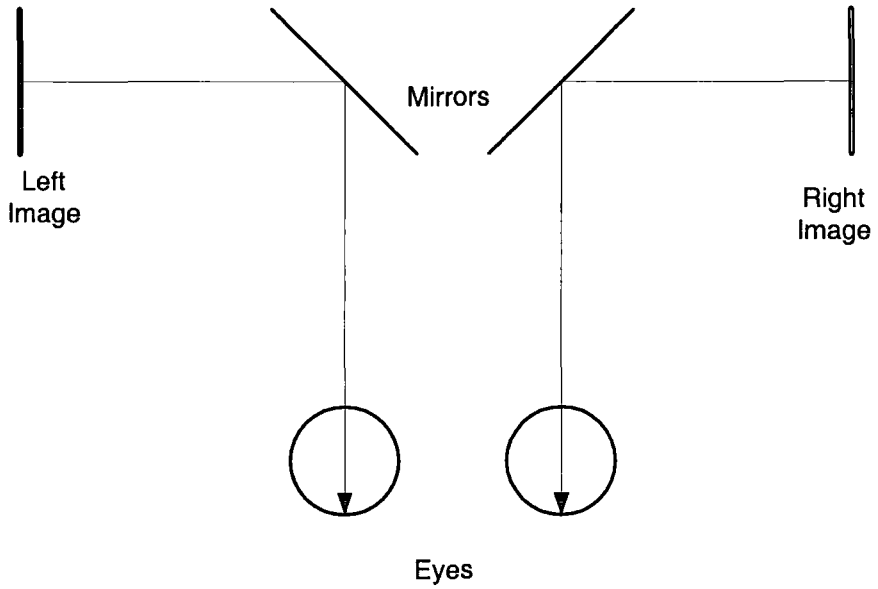


Figure 2.5: Stereoscope Design

2.3 Stereo Display Hardware

2.3.1 Stereoscopes

Wheatstone [112] was the first researcher to publish a formal paper on the process of stereopsis back in the 19th century. In his paper he gave detailed discussions of phenomena involving binocular viewing. He identified that stereopsis occurred due to humans having two eyes separated by an interocular distance and used this knowledge to make the first *stereoscope*, a device for artificially simulating stereo viewing. The principles he identified for constructing stereoscopes are the same as used in the making of modern day stereo display devices.

Wheatstone discovered that for a stereoscope to operate it is necessary for the device to project a left and right image into the respective eyes. The projected pair should have a level of horizontal disparity between the them. This simple design is depicted in Figure 2.5.

2.3.2 Modern Stereo Displays

Modern stereo display systems aim to deliver an observer with additional cues present in binocular viewing as well as the normal monocular cues present on traditional 2 dimensional display systems. Hence, the displays attempt to bring a more realistic stereo experience to observers similar to what observers would experience in normal non-virtual viewing. As discussed in the preceding sections of this chapter the overall perception of depth in a scene is not solely reliant on binocular vision. In fact, even in the presence of no binocular visual information the human visual system is able to decipher large amounts of depth information. This means that individuals with little or no binocular vision are still able to perceive depth. However, the presence of binocular disparity gives additional depth cues which can aid an observer in establishing an accurate perception of the depth encoded in observed scenes. Hence, a display that can provide the ability to simulate an accurate stereo representation of a given scene could increase the realism of viewing experiences. Such displays are known as stereoscopic display devices.

Wartell [110] states that stereoscopic displays can be placed into two main categories: *Volumetric 3D displays* and *Surface 3D displays.*

Volumetric displays [65][6] illuminate true locations within 3D space. The display projects the light into the 3D space around the display panel. This means that the display can be viewed at most angles to the display surface as well as the ability to support a *multi-view experience* - multiple observers using the system simultaneously. Volumetric displays include the subcategories of holographic displays, swept volume technique displays and static volume technique displays.

Surface 3D displays emit light from a single surface source to give an illusion of depth. The display is used to project the left and right image of the stereo pair into the appropriate observer's eyes. This means that is it necessary for an observer sit in an appropriate position in relation to the display.

Stereoscopic displays are a subset of the Surface 3D displays category. The use of stereoscopic display systems appears to be the future of stereo viewing and most manufacturers of stereo displays are focusing their research into this area of stereo display hardware development.

Stereoscopic display devices are constructed using CRTs or LCD panels. The display does not physically project the image into the 3D visual space around the display instead it uses a mechanism to project the appropriate images into the left and right eyes of the observer. Stereoscopic display devices can be categorised further into *autostereoscopic* and *stereoscopic* displays [42].

Stereoscopic Displays

Stereoscopic displays require observers to wear a device such as a headset or goggles to ensure that the correct image is relayed to the appropriate eyes. There are two subcategories of stereoscopic display systems:

1. Helmet devices - These helmets contain small LCD or CRT monitors for the left and right eye. The images are presented directly into the observer's eyes.

2. Glasses or goggles - The glasses or goggles are usually LCD shutter glasses or polarisation glasses. LCD shutter glasses function by presenting a left and right view in rapid succession giving the stimulation of depth similar to the ordinary binocular disparity cue encountered. Polarisation glasses perform the same functionality but are usually used in conjunction with a projection display and effectively provide a rapid succession of the viewed scene on the external display.

Autostereoscopic Display Devices

> 'Autostereoscopic displays present a spatial image to the viewer without the use of glasses, goggles or other visual aids.' [33]

There are various autostereoscopic displays in today's market of varying degrees of projection quality. Some of these displays include the ability to provide observer tracking [117]; multiple viewing regions [19] and 2D to 3D switchability.

Autostereoscopic displays [118][42] provide users with a stereo viewing experience without requiring any additional equipment. The display projects both the left and right images and separates each view for the intended eye. Hence, they are perceived to be less intrusive to the observer than stereoscopic goggles or glasses. Autostereoscopic displays are constructed using technologies such as Parallax Barrier, Lenticular Optics and Micropolarisers. These devices are present in the display panel and are used to only allow light from a stereo image to pass into the correct eye.

Parallax Barriers

The structure of a parallax barrier display [33][118][41] consists of an array of optical apertures in front of the display panel known as a parallax barrier. The apertures used in the construction of the parallax barrier are made of a highly robust and anti-reflective substance. The apertures enable light to be projected into the intended eye and prevent it passing into the other eye. The parallax barrier is aligned perpendicularly in front of the display panel with the display panel itself consisting of a series of interlaced images corresponding to the left and right views of the stereo pair. When the slits of the parallax barrier are correctly aligned they act as a windows onto the display panel. Hence, a correctly positioned observer should be able to see the right image projected into their right eye and left image into their left eye where both images originate from the display's surface. The general viewing position for this kind of display is directly in front, central and orthogonal to the display surface.

The size of the slits of the parallax barrier can be altered to allow for a multiview system. Increasing the size of the slits can be used to create a greater number of views by using the principle of diffraction. The display can be used to project either the same, or different viewpoints of the displayed scene

through each of the display's possible projections zones. This would result in the ability of the display to cater for multiple users and/or given a sense of 'look-around' when an observer moves parallel to the display plane. This may give a greater realism to the viewing experience.

Parallax barrier suffer from a number of disadvantages due to their design. These include the following:

1. Use of the parallax barrier to clock unintended views also results in the blocking of light from the display panel. Hence, it is necessary for the display to implement a bright-diffused light source to be located behind the display surface to enable the luminosity of images to be maintained.

2. The displays rely on the spatial and directional information to be spatially multiplexed on the display (the interlacing of the left and right image) and this design feature leads to the following problems:

   - Imprecise viewing positions can result in the observer viewing the image *pseudoscopically.* Pseudoscopical refers to a phenomenon where the left and right images are observed reversed (the left image is observed by the right eye and the right image by the left eye) and hence the depth can be flipped inside out.
   - The resolution of the display limits the number of possible viewing projections which can be comfortably observed on the display without losing too much resolution. Regardless, the projection of a stereo image will reduce the resolution of the display by half since there will be a minimum of two *viewing windows.*

3. Utilising a parallax barrier grating can result in problems caused by diffraction [41]. Wider apertures result in brighter images and reduce diffraction problems, however, they also cause a reduction in performance by producing poorly defined viewing windows. Viewing windows are a volume where the observer's eyes must be positioned to observe the images correctly. A narrower aperture results in a reduction in the bright-

ness of the images but with a more accurate window definition; however, a narrowing aperture results in more diffraction which can affect the quality of the viewing windows produced. Hence, it is necessary to have an appropriate balance for the size of the aperture implemented by the parallax barrier.

Lenticular Displays (Single LCD Panel)

Lenticular displays [41][33][65] are similar in structure to parallax barrier displays, however they differ in their mechanism for splitting the left and right images. Lenticular displays perform the image separation by the use of vertically arranged cylindrical lenses. The lenses are arranged in an array where each lens focuses the interleaved stereo image located on the display surface into the appropriate observer's eyes. The display containing the images to be projected by the display is known as a *lenticular panoramogram.* This is located on the display panel behind the lens. Since the lenticular lens sheet is able to transmit all of the projected light this results in the display being more efficient and brighter than a parallax barrier. The lenticular sheets are moulded from high quality plastic.

The major disadvantages of using a lenticular display systems include the following:

1. The production of the high quality uniform lenticular lens sheets is expensive.

2. Achieving correct alignment of the lenses with the lenticular paroramogram is a difficult task which is necessary to obtain correctly positioned projections and distortion free stereo images.

3. The lenses tend to suffer from scatter and poor antireflection performance which can cause a degrading in the viewing experience.

2.3.3 Human Factors and Autostereoscopic Displays

Autostereoscopic displays suffer from numerous problems that can cause degrading of the viewing experience. These are in addition to problems created due to the display's construction. These problems [120][59][22][41] include crosstalk (ghosting), inter-channel variations in terms of brightness and contrast, display refresh rate, inadequate display bandwidth (resulting in a decrease in image and depth resolution), channel misalignments and image misalignment.

These problems can degrade the viewing experience for observers. They usually cause a reduction in the ability of the observer to fuse the stereo image. Ultimately, this results in a reduction in the quality of the stereo experience. Usually, image degrading is more pronounced at image regions with high contrast levels and hence apparent at feature edges [76] and with increased scene encoded depth [42]. Hence, limiting each of these possible quality disturbances is paramount in achieving an overall high quality stereo viewing experience.

Scenes observed on displays have less resolution than scenes viewed naturally. This means that less texture is provided affecting the texture perspective interpretations and thus the overall depth perception. Additionally, few displays can provide the range of luminance and contrast that are typically experienced in the real word and hence since stereo acuity (the ability to resolve images) is dependent on brightness this can be reduced.

Pastoor [77] discusses the results of experiments performed to determine criteria for evaluating and providing solutions for stereo display systems based on the principles of human perception. Pastoor discusses many quality issues in the context of stereoscopic stereo displays including:

- **Minimum Display Size**
  Visual size distortion can occur due to the use of small display screens. This causes the retinal size of the observed objects to be different from

the sizes expected from real world viewing conditions. This is due to perceptual distortions. The result is a reduction in the viewing comfort when depicting disparities similar to those experienced in the real world.

- **Exploitable Depth Range**
  Observations and experimentation have found that large binocular disparities projected on stereo displays tend to cause adverse side effects. However, in real world viewing experiences the human visual system is able to cope with an extensive range of disparity levels. It is unknown as to the precise reason why the visual system struggles with virtual stereo but it is advisable to limit the depth ranges available to prevent adverse viewing conditions.

- **Flipping in Multiview Displays**
  Multiview stereoscopic display systems can operate such that when an observer changes their head position the perspective of the viewpoint changes. Studies have shown that this change of view can cause noticeable jumps of the image from one perspective to the next. A large number of smoothly projected views are required to make this image flipping unnoticeable. If these multiple views are not available then the result can be a scene which can be confusing to the moving observer.

- **Crosstalk**
  Crosstalk [33][61] is a problem encountered in stereoscopic display systems. It occurs when an image of the incorrect perspective is observed by the eyes; that is part of the image intend for the left eye is seen by the right and/or visa versa. Crosstalk is caused by imperfections within the arrangement of the physical display hardware. In the current autostereoscopic display market, display systems exhibit cross talk usually in the range of 1-10%.

  Ghosting is the perceived crosstalk when images are viewed on stereo displays. Ghosting varies with the brightness, contrast, resolution and parallax of the images. Images with large amounts of parallax and high

levels of contrast cause a greater degree of ghosting when viewed on a display.

Visual impairments mean that it is not possible for all observers to use autostereoscopic display systems. Obviously, users with defects in their binocular vision will not be able to successfully use such displays. Similarly for users with only partial binocular vision, disadvantages present in the use of the displays may outweigh any benefits obtained and hence autostereoscopic displays may not be suitable. Additionally, observers who have unresolved defects in their refractive vision may find problems with focusing the images on the display systems. Evidence has suggested that even those observers with small defects in their ability to focus images can have problems when viewing images specifically with the perception of disparity ranges depicted in the images.

2.4 Binocular Disparity on an Autostereoscopic Display

Figure 2.6 demonstrates a basic mathematical model of disparity on an autostereoscopic display, where an observer sits orthogonal to the screen as demonstrated in various examples in the literature [41][111][80][38].

Using Figure 2.6 it is possible to derive formulae for each of the variables when considering the following constants, assuming a given observer and display system:

- i - the interocular distance (eye separation).
- D - the screen to observer distance.
- p - the pixel width of the display.

In summary, the equations calculated in this thesis using the display model depicted in Figure 2.6 are given in Equations 2.1, 2.2, 2.3, 2.4, 2.5, 2.6 and

2.7. For uncrossed disparity, it is assumed that the eyes can only converge onto a point and hence parallel viewing and divergence are not considered. In addition, to above, the following variables are defined and used in the given equations:

- n - number of pixels.
- s - screen parallax.
- XRight - The x coordinate of the screen intersection as observed by the right eye. The coordinate 0 is taken for the left side of the display.
- XLeft - The x coordinate of the screen intersection observed by the left eye. The coordinate 0 is taken for the left side of the display.
- V - The distance from the observer to the point of vergence (perceived depth).
- Sd - The distance from the display to the point of vergence.
- VD - Vergence Difference.

$$s = XRight - XLeft \tag{2.1}$$

$$s = p * n \tag{2.2}$$

$$V = \frac{iD}{i - s} \tag{2.3}$$

$$Sd = D + \frac{iD}{s - i} \tag{2.4}$$

$$VD = 2arctan(\frac{i}{2D}) - 2arctan(\frac{(i - s)}{2D}) \tag{2.5}$$

$$Crossed : VD = b - a \tag{2.6}$$

$$Uncrossed : VD = b - c \tag{2.7}$$

The derivation and an explanation of some of these equations is given in Appendix A. In addition to this the following should be noted:

Vergence difference is a special case of angular disparity. In addition to the difference between the angles of two points in space, vergence difference is defined to be the difference between the point of convergence and the point of accommodation. Vergence difference is sometimes referred to as 'differential parallax' [98] or 'phoria' [107].

The point of focus, on autostereoscopic displays, is assumed to be on the screen plane at the point of zero disparity [65]. Figure 2.7 demonstrates vergence difference, for both crossed and uncrossed disparity objects in relation to the screen plane. For crossed disparity $VD < 0$ and is often referred to as *crossed-vergence*. For uncrossed disparity $VD > 0$ and is often referred to as *uncrossed-vergence* [110].

Hodge [39] and Wartell [110] stated that the *Horizontal Vergence Angle* (HVA) is equivalent to vergence difference. They use these two terms interchangeably. However, calculations performed in this thesis determined this to be incorrect. The HVA and vergence difference in most image situations will be a similar value, however, they are not mathematically equivalent. In Appendix B a discussion and a mathematical derivation of the inequalities of these two variables is provided. For this reason the use of HVA is discarded within this thesis.

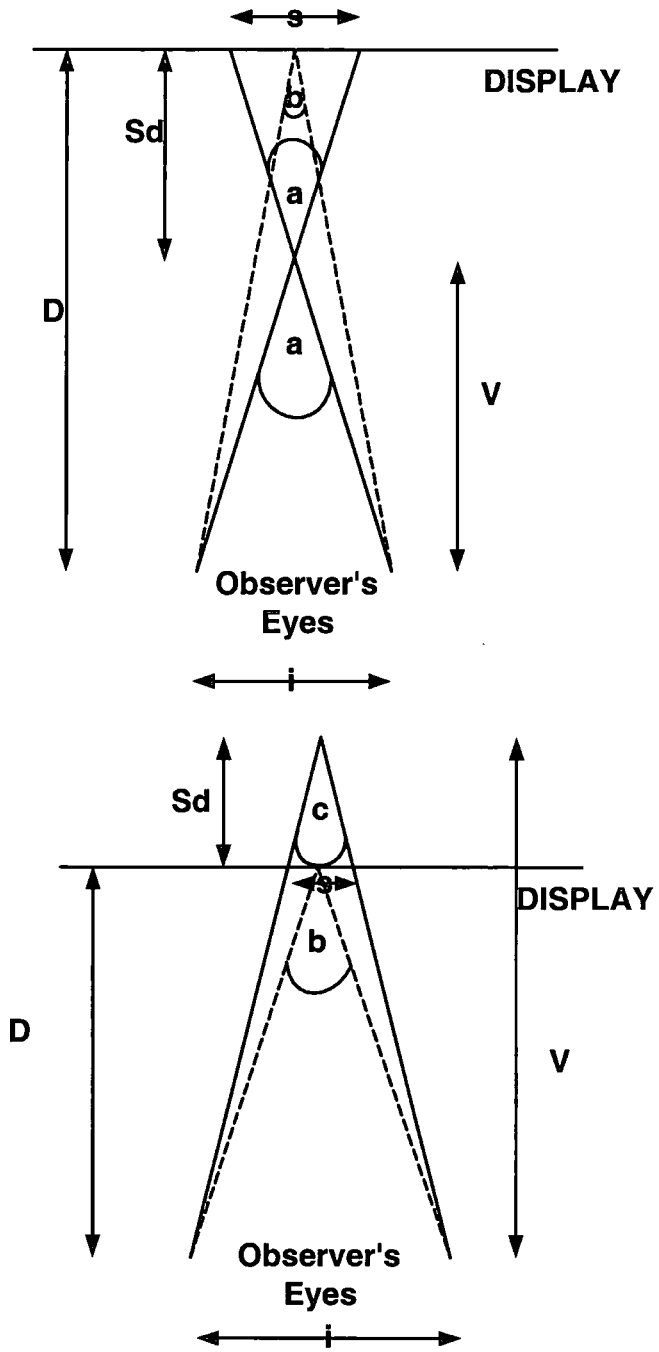


Figure 2.6: Model of both Crossed (top) and Uncrossed (bottom) Disparity

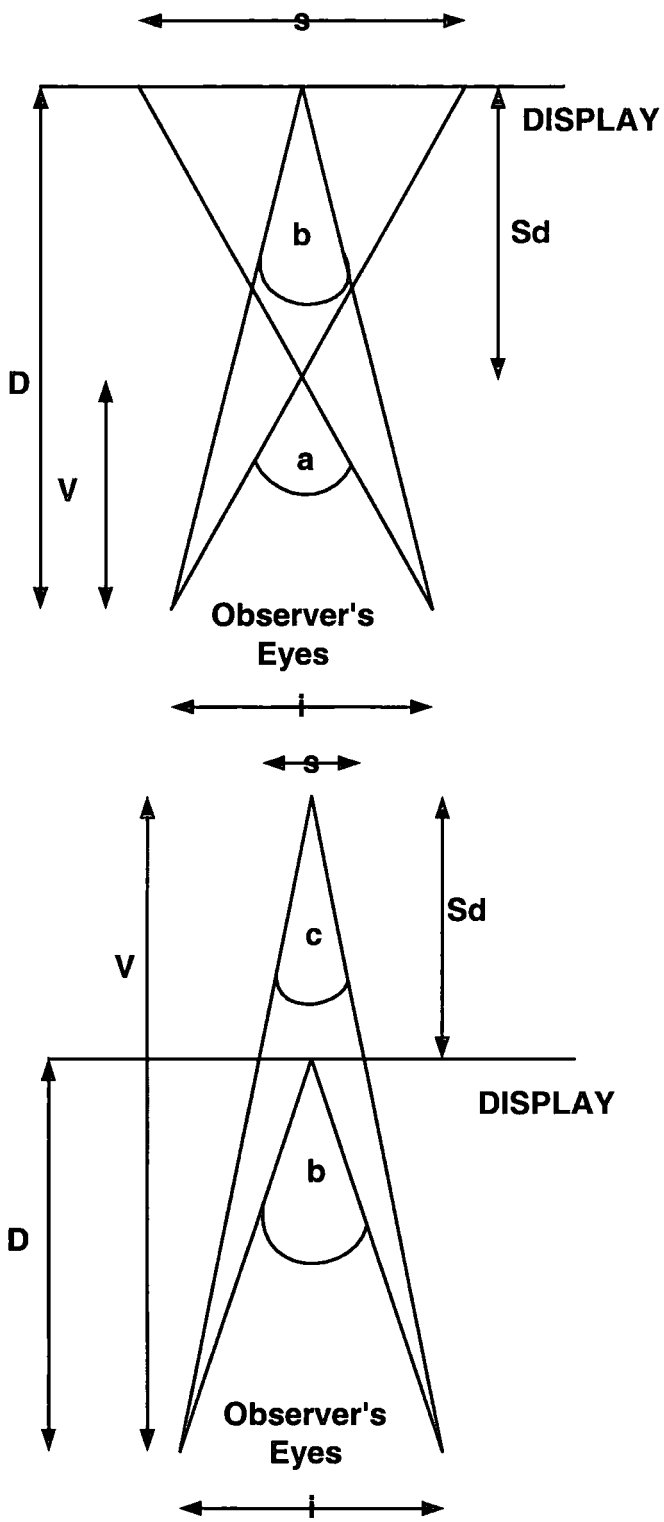


Figure 2.7: Model of Focus and Vergence Angles with Crossed (top) and Uncrossed (bottom) Disparity

Chapter 3

Stereo Image Quality and Human Factors

This chapter will initially provide an introduction to the currently available stereo image quality systems and provide a brief analysis of them. An in-depth discussion of the human factor limits identified within human vision psychology research will be made and related to autostereoscopic display systems. Other identified influences on stereo image quality are concisely summarised in the concluding sections of the chapter.

This chapter will be used as a basis for the formulation of a model for the development of a system to assess the quality of stereo images.

3.1 Stereo Image Capture

The viewing quality of stereo images is determined both by the quality of the display projection system and the observed stereo image. Errors present in either the projection system or the observed images can result in adverse side effects including: nausea, eye strain, diopia, blurry vision and headaches [111][65][41][15]. This means it is necessary to consider both the display hardware and driving software; and the stereo image to be observed in order to provide a high quality stereoscopic viewing experience.

The phrase 'stereo image quality' in terms of stereo images is poorly understood. There are no definitions of what results in the quality of stereo images viewing experiences degrading and how such aspects may cause adverse side effects. The term quality in reference to stereo images is used by professionals in the area who observe stereo images and from their experience assess if they are suitable for viewing or if they contain problems. However, there is no quantifiable or qualitative definition which states when a stereo image becomes non-viewable or what aspects of the image make it so. This means that it is difficult for non-experts to assess stereo images and decide if they are suitable for viewing.

One method for establishing a generally consistent level of quality of stereo images is to implement controlled procedures in their capture.

3.1.1 Stereo Image Generation

To limit the possible problems associated with observing stereo images it is possible to use established stereo image generation techniques or specific stereo cameras to ensure that stereo images are captured as error free as current research permits.

Stereo Camera Hardware

There are a few purpose built stereo cameras available. The majority of these are aging by today's standards and include the following:

- **Stereo Lens Adaptors**
  Stereo lens adaptors [53][79] are camera lenses available for standard SLR cameras to enable them to capture stereo photographs. The lens adaptors provide the ability to capture two images simultaneously of the same scene with a level of horizontal disparity between them. Since the stereo lens adaptors usually use mirrors to capture two images concurrently the conditions for the image capture remain the same for each image of the stereo pair.

- **Stereo Cameras**
  Current available stereo cameras [79] are similar to ordinary cameras but as opposed to one lens they contain two fixed lenses on the camera body. A set horizontal distance separates the lenses to allow horizontal disparity to be present in the capture of the stereo image.

Both these hardware methods for capturing stereo images produce questionable quality in the final stereo image capture.

Techniques for Stereo Image Capture

Kohno [55] discusses a very simplistic method for capturing stereo images. His proposed method involves using two single vision cameras and taping them together so that the distance between the two image captures is fixed. This method fixes the horizontal shift regardless of the depth present in the captured scene.

Discussed within [48][44][55] is another process for capturing stereo images. The discussed method involves capturing the left image of the pair and then shifting the camera horizontally a prescribed distance to capture the right image. This capture technique is based on the fact that the horizontal movement

of the camera causes the left and right images to have an overlapping region in addition to areas of the images that do not overlap. The additional non-overlapping sections are known as the *parallax* [44]. This parallax should be minimised to provide good stereoscopic images; that is, the horizontal disparity present in the image should be limited.

Theories [48][68] regarding the capturing of stereo images state that the optimal distance of the camera separation for the capture of stereo images is dependant on the distance from the observer to the scene. For captured scenes closer than 2m to the camera, a camera separation of less than the interocular distance should be used. For scenes in excess of 2m a larger camera separation distance than the interocular distance is required. This contradicts other theories which state that the horizontal disparity should be limited to the interocular distance.

Stereo pairs with a horizontal shift greater than the interocular separation are referred to as *Hyperstereo pairs* and less than the interocular distance are referred to as *Hypostereo pairs* [68]. Another theory of stereo image capture is known as the 1/30th rule [68] which states that the camera horizontal shift should be 1/30th of the distance to the nearest object within the captured scene.

3.2 Stereo Image Quality Systems

Computer vision is a multiple disciplinary area bringing together aspects of image processing, artificial intelligence, human psychology, human physiology, philosophy and computer graphics [83]. Computer vision is a branch of image processing concerned with the computer processing of real world images. Typically it involves low-level image processing to enhance images and high level processing to discover patterns and aspects in images that were previously unknown. Low-level processes include noise removal and increasing the contrast within images. High level processes include pattern recognition and

image comprehension to recognise features and objects in presented scenes [51].

The research area involving determining the quality of stereo images is relatively new. Extensive research has been performed in the determination of quality metrics for 2 dimensional images and although this can be applied to the individual images in stereo pairs, the quality attributes of the actual stereo pair remains poorly understood.

A summary of the current research regarding 2 dimensional image quality is summarised in Wang et al. [105] and Chalmers et al. [12]. Discussed within [106][81][105] are the different forms of image quality assessments for 2 dimensional pictorial images. They state that the factors affecting image quality include contrast ratio, resolution, brightness and compression rate. An examination of 2 dimensional image quality assessment is beyond the scope of this thesis.

3.2.1 Stereo Image Quality

Assessing the Accuracy of Correspondence

Stereo images capture a scene with different levels of disparity present throughout. This is due to the different levels of depth present in the captured scene. As discussed in Chapter 2, the depth is depicted in the images as differences between corresponding pixels in the left and right images. This difference is demonstrated in Figure 3.1. The diagram depicts two pixels which are matching between the left and right images but are present at different horizontal coordinates i.e. there is a level of horizontal disparity between them.

The process of analysing a stereo image and determining the corresponding pixels for each pixel in the left and/or right image is known as correspondence. The process of establishing correspondence appears on initial inspection to be a simple procedure since the human visual system is able to perform this process without any need for conscious human interaction. To perform the task of cor-

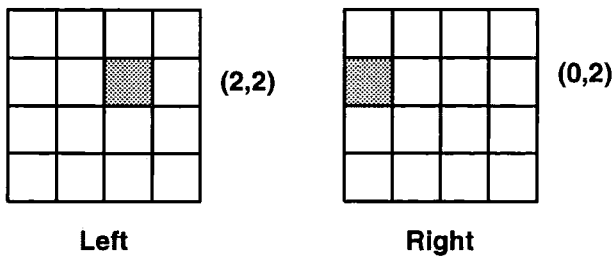


Figure 3.1: Pixel Correspondence

respondence on a computer involves the production of accurate algorithms or heuristics. Current research has not discovered an absolute solution to the correspondence problem instead there only exists heuristic solutions.

Many of the developed heuristics are highly accurate in performing correspondence; however, they are restricted in their ability and can only successfully analyse a subset of stereo images where certain conditions are adhered. Examples of these types of solution are given by Olague et al. [73]; and Bleyer and Gelautz [8].

Other stereo correspondence solutions are able to analyse the majority of stereo images, however, they produce a solution which may contain a large proportion of errors. Examples of these solutions are StereoMatch [91] and TreeDP [99].

Usually stereo correspondence systems stipulate certain image criteria. The majority of stereo correspondence heuristics stipulate that the examined images are *rectified*. Rectification is a process where stereo images are horizontally aligned and all vertical disparity is removed. Often stereo images become unaligned vertically when they are captured. This is especially the case when the image pair is captured without a tripod and a stable horizontal shifter because even the slightest motion can introduce vertical errors.

The majority of correspondence algorithms perform searches using the same vertical coordinates in the images. This means that it is necessary for the

images be vertically aligned otherwise the heuristics will provide accurate solutions. Various rectification algorithms and systems exist including Fusiello et al. [27] and Chen et al. [13].

Standard correspondence heuristics perform correspondence by determining matches between the pixels of a stereo image. They attempt to find a matching pixel by searching a designated search space for pixels exhibiting the best match. The match criteria is generally heuristic specific. The formulation of the search space area can be derived using various techniques. One such solution by Kawai et al. [52] uses edge detection to produce a set of boundaries in the images which can then be used to produce image segments. This image sectioning results in limiting the regions for the search for correspondence. Other solutions merely stipulate a horizontal range in which the matching pixel should reside.

Testing Correspondence

Scharstein and Szeliski [89] created a set of test stereo images for use in correspondence systems evaluation. Other researchers use these images as a framework for testing their own stereo correspondence and disparity map generation algorithms. The results of these tests involve evaluating how the generated disparity map compares to the groundtruth image. This effectively allows algorithms to be compared.

Stereo Quality Assessment

Egnal et al. [21] provides one of the most in-depth discussions regarding the assessment of stereo correspondence quality. He summarises 5 metrics including: Single View Stereo (SVS), L/R Consistency, Matching Score Metric (MSM), The Curvature Metric (CUR) and Peak Ration (PKR). The most comprehensive and commonly used metric is LRC.

LRC involves a stereo image being processed so that for each point in the left image A correlation is made with the predicted corresponding point in the right

image. The calculated right point is then used to calculate the corresponding point in the left image. The difference in the original and correlated pixel position can be used to assess the predicted quality of the correspondence. Figure 3.2 demonstrates this procedure.

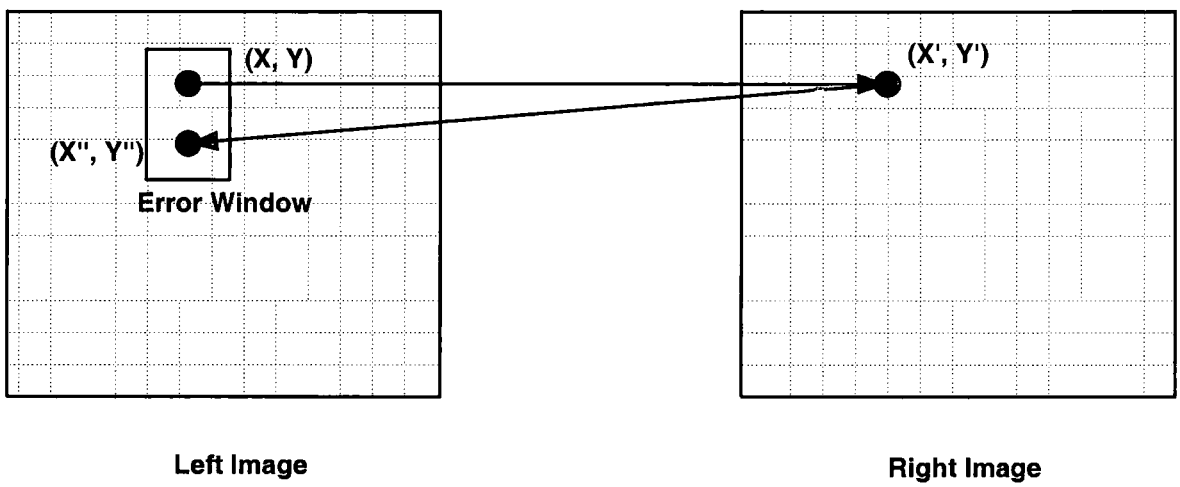


Figure 3.2: L/R Consistency

If $(X, Y) = (X'', Y'')$ it is assumed that there is a prefect correlation within the image. An error window can also be introduced within the procedure to allow for imperfect matches. This can be used to account for errors due to integer interpolation round up and a small amount of noise in the images. Correspondence correlations lying outside of the window are assessed to be of unsuitable quality. This method can be used to highlight possible image problems including noise and contrast. However the analysis assumes that the image correspondence is accurate and does not consider the errors present due to invalid correspondence. Additionally, the metric does not consider the quality aspects identified via human factor analysis.

Examples of other correspondence metrics and assessments are given within [21][96][54]. However, the literature is very limited in its discussions of stereo image quality in terms of its definition and implementation. Discussed within Kinder Gentler Stereo [93] is a paradigm for stereo image capture, rendering

Figure 3.3: Disparity Map generated via StereoMatch for Middlebury Image - Cones

and displaying stereo images. It intends to provide observers viewing experiences without cue conflicts, eye wear and viewing zones. Kinder Stereo effectively searches for a stereo viewing experience that is free from any conflicts, however, it does not state what this would entail or how it would be measured. Its aspirations seem rather idealistic in terms of current knowledge.

3.3 Disparity Maps

Disparity maps represent the encoded horizontal disparity in stereo images and are a graphical means to represent correspondence by using an intensity level representation to denote the difference between the two images. To interpret a disparity map for use on an autostereoscopic display it is necessary to possess a translation scale for intensity level to screen parallax (horizontal disparity).

Usually disparity maps are constructed based on either the left of right image of the stereo pair. A disparity map derived from the left image will feature pixels at the same coordinates as the left image with the intensity level in the disparity map representing the amount of disparity present between them. The same is true for those disparity maps derived from the right image of a stereo pair. Using the intensity level derived from the disparity map and translating this into a quantitative pixel difference means that it is possible to determine the correspondence between the left and right image. Figure 3.3 demonstrates a disparity map generated from a pair of stereo images.

A disparity map can be generated either via groundtruth methods or as a product of heuristic analysis of the images. Obtaining a groundtruth disparity map involves a complex procedure that is both time consuming, requires specialist equipment and a large degree of skill to complete successfully. A groundtruth disparity map represents the actual disparity encoded in a stereo pair - or as close as current knowledge permits. They are generated at the time of scene capture and are currently the most accurate demonstration of a stereo image's encoding of disparity. For this reason groundtruth disparity maps are used as a benchmark for the assessment of generated disparity maps and correspondence algorithms. Due to the complexity and costs involved in their production very few stereo images have a groundtruth disparity map associated with them. The core set of images provided with groundtruth disparity maps are located at Middlebury [86] and are utilised extensively within the literature and this thesis.

Disparity map generation algorithms like those of correspondence solutions are limited in their ability to produce accurate results. Research into disparity map solutions has not at the time of writing produced a perfect solution. The production of disparity maps generally involves the formulation of a heuristic to present a reasonable solution to the correspondence problem between a stereo pair. Since there is no perfect method of providing a 'perfect' stereo correspondence solution this results in such methods of disparity map generation being limited in their accuracy.

Due to the inaccuracies in producing disparity maps it follows that generated disparity maps are less accurate than groundtruth maps. An extensive summary of disparity map generation algorithms is given in [87][88].

3.4 Quality Issues in Stereo Images - Human Factors Analsyis

Quality issues associated with stereo images can cause stereo viewing problems even when observed on a theoretically perfect display system. Hence, it is necessary to produce stereo images using tightly controlled procedures. The problems found specifically in stereo images which can affect their viewing quality include:

- Differences in the individual image sizes.
- Differences in the individual image resolution or file formats.
- Stereo image compression [113].
- The encoded disparity levels [98][119][23][60].
- The randomness of the encoded disparity [9].
- The presences of frame cancellation [98].
- The corresponding area intensity level differences.
- The corresponding area colour difference.
- The corresponding area contrast level differences [43][115].
- Image rectification[24].

Some of these problems exist in normal stereo viewing; however, some are specific to the viewing of stereo as a virtual experience. This thesis will examine these identified problem areas and summarises the current research knowledge within each of the areas in the context of autostereoscopic display systems.

3.4.1 Controlling Disparity Levels

Discussed in Chapter 2 is the utilisation by the brain of the difference in the position of the fixation point and points not on the horopter to formulate binocular disparity information. This disparity causes an offset on both eyes' retinas corresponding to the disparity between the two points. Diopia occurs when there is too much disparity between the two retinal images [50] this results in the observer seeing two images of the visual scene simultaneously (summation).

The visual system is able to interpret disparity only up to a threshold without diopia occuring. To prevent this the disparity must be limited to reside within Panum's Area. Panum's area is discussed by Howard [43] and Ware [108].

The use of autostereoscopic displays introduces additional viewing problems infrequently encountered in normal viewing. One of the introduced problems includes the conflict between the occulomotor cues (accommodation and convergence) [92][16][108]. When observing a stereo image on an autostereoscopic display system the observer accommodates on the originating source of the projection; the screen plane. The display causes an implied virtual image either in front or behind the display screen plane. The eyes converge onto this point in space. This results in a conflict between the cues interpreted by accommodation and convergence within the brain. To reduce the conflict between the oculomotor cues it is necessary to limit the difference in the reported accommodation and convergence points. Lipton [61] discusses limiting the disparity within images to reduce the accommodation and convergences conflict in order to improve overall image quality and viewing comfort.

The limitation of comfortable observable disparity in images is not just restricted to the problems caused by the accommodation and convergence conflict. In general, if the amount of disparity exceeds limits imposed on the visual system the result is the inability to fuse images and hence diopia and other adverse side effects.

Additionally, the human visual system requires a minimum amount of disparity to be present before binocular disparity is detected. The minimum amount of disparity required to bring about depth perception is known as the *stereo acuity threshold.* Stereo acuity is also a measure of the smallest perceivable change in angular disparity between two objects. Stereo acuity is approximately 2-60 arc seconds [41][103] when observing stationary objects with normal binocular vision. The average observer's personal stereo acuity is approximately 20 arc seconds [41].

3.4.2 Disparity Limits

Discussed in Chapter 2 is a model of disparity on autostereoscopic displays. Observations and experimentation have discovered that the use of autostereoscopic displays can cause observers numerous side effects. These side effects can be increased, or become more apparent, when horizontal disparities are large, or in the presence of vertical disparity [45]. Excess horizontal disparity in images can cause headaches, eyestrain, fatigue, disturbances of vision and nausea etc. [111][65][41]. Hence, it is beneficial to limit the amount of disparity in stereo images to prevent possible adverse side effects to the observer. Detecting levels of disparity above the prescribed limits can be used as an indicator for an image's suitability for stereoscopic viewing.

The literature contains numerous references to prescribed limits on both crossed and uncrossed disparity, for the comfortable viewing of stereo images. These limits are derived from human factors research and may be used to place a limit on the amount of depth displayed both in front and behind the display panel of an autostereoscopic display.

The examined limits within this thesis are taken from various sources [92] [1][121][23][98][116][48][78][119][60] and will be summarised in the following sections of this thesis. Other limits can be found in [39][40][94][100] but will not be considered. Each of the limits provides a suggestion to an acceptable

limit on the range of comfortable disparity, for a normal observer, with functional binocular vision. They provide a summary of the diverse human factors recommendations for disparity perception.

3.4.3 A Practical Calculation of the Disparity Limits

To illustrate the behaviour of the limits a summary has been made of the limits applied to a 15" autostereoscopic display with a resolution of 1024*768, a horizontal pixel size of 0.29766mm and with an acceptable viewing distance of 700mm. Such displays are DTI (2015XLS and 2015XLV), VRex (15fp), and Pavonine (PA-3D15CV and PA-3D15GL). The model of viewing using an autostereoscopic display provided in Chapter 2 is utilised and a table of the resulting calculations can be found in Appendix C - Table C.1.

Siegel and Nagata [92] - Microstereopsis: A Limit on Screen Parallax

Siegel and Nagata [92] stated limits on microstereopsis of 1mm of screen parallax, stating this level of disparity can be detected by two-thirds of observers. Microstereopsis describes the smallest recognisable or fusible disparity in humans. The results of experimentation show that small disparities stimulate accurate and pleasant depth sensations and also cause a reduction in the adverse side effects associated with stereoscopic viewing [93]. Hence, these should be considered as an alternative to the use of large disparities. On the examined displays, the minimum permitted screen parallax is 0.29766mm; which results from one pixel of screen parallax. The 1mm limit proposed by Siegel and Nagata equates to 3.560 pixels of screen parallax.

Akka [1] - Screen Parallax Limit

Akka [1] states a maximum limit on the permissible screen parallax of a 3.5% of the display screen width. For the examined 15"displays the screen width is $304.8mm$. Hence, this limit denotes a maximum screen parallax for crossed and uncrossed disparity of $10.668mm$. This translates into a screen parallax

of crossed and uncrossed disparity equal to $35.907 pixels$.

Yeh and Silverstein [121] - Vergence Difference Limit

Yeh and Silverstein [121] state that vergence difference should not exceed $27 minarc$ for crossed and $24 minarc$ for uncrossed disparity, otherwise diopia will occur. This limit is based on an image exposure time of $< 200 ms$. This time constraint means that eye vergence onto the projected point will not occur fully and hence depth perception will be limited.

Yeh and Silverstein [121] - Vergence Difference Limit

Yeh and Silverstein [121] state that the vergence difference should not exceed 4.93^o for crossed and 1.53^o for uncrossed disparity, otherwise diopia will occur. These limits are based on an image exposure time of $> 2s$ allowing time for vergence to fully occur.

Farrell and Booth [23] - Projection Depth from Screen Limit

Farrell and Booth [23] stated that the projected depth from the screen should be limited to $+/- 0.75 Diopters$.

Definition of a Diopter

A discussion is made within [120] regarding the definition of a diopter. Farrell and Booth [23] state that the disparity should not exceed $0.75L$ of the screen to observer displacement. Screen diopter distance is defined to be [120]:

> 'The optical viewing distance (in diopters) of the screen is obtained by obtaining screen distance from the eyes in meters, then dividing that into 1 to get the distance in diopters.'

This results in Equations 3.1 and 3.2, where D is the distance of the observer to the screen plane and Dio is the distance measurement:

$$\text{Diopters} = 1/\,\text{Dio} \tag{3.1}$$

$$Sd = D - 1/Diopters \tag{3.2}$$

For a display with a viewing distance of $700mm$ this equates to $1/0.7 = 1.42857D$. Performing $+0.75D$ (as denoted by Farrell and Booth's limit) gives a perceived depth of 2.1787D. This equates to $1/2.1787 = 459mm$ which is a depth relative to the screen plane of 241mm. Performing -0.75D gives the permitted uncrossed disparity, equating to a perceived depth of 0.67857D. $1/0.67857 = 1473mm$ which is a depth relative to the screen plane of 774mm.

Valyus [98] - Vergence Difference Limit

Valyus [98] stated that the vergence difference should not exceed 1.6^{o} ($96minarc$) for both uncrossed and crossed disparity, otherwise diopia will result.

Valyus [98] - Screen Parallax Limit

Valyus [98] stated that the screen disparity should not exceed 3% of the observer to display distance. For the examined displays the observer to screen distance (D) is 700m. This equates to a limit on the permitted screen parallax of 21mm or 70.550 pixels.

Williams and Parrish [116] - Projection Depth from Screen Limit

Williams and Parrish [116] stated a limit of -25% - +60% of the observer to screen distance for the projected depth from the screen plane (Sd). On the examined displays with $D = 700mm$ this equates to a value of $Sd = 175mm$ crossed and $Sd = 420mm$ uncrossed projected depth disparity.

Patterson and Martin [78] - Vergence Difference Limit

Patterson and Martin [78] state that at the fovea, the maximum disparity before diopia occurs is $1/10^o$. As discussed in Chapter 2, the fovea is located on the retina of the eye. It is an area primarily populated by cones, which allows for high visual acuity. Visual acuity is a measure of the spatial resolving capacity of the eye; hence, something termed to have high visual acuity is able to resolve fine detail within an observed scene. The fovea itself is only $2 - 3^o$ in size [101] but contains a point residing approximately central where images caused by the fixation of the eyes are located. Small deviations from the fixation point located on the fovea, of as little as 5 minarc, can cause considerable loss of visual acuity [56].

The limit of $1/10^o$ from Patterson and Martin refers to experimentation for objects viewed on the fovea and hence should be viewed as a worst case scenario.

Hodge and Davis[40] produced a mathematical model of stereoscopic viewing. Using this model and the calculations, Hodge and Davis determined that retinal disparity is mathematically equivalent to angular disparity. Using this equivalence, it is possible to translate the limit given by Patterson and Martin into vergence difference. The result is that vergence difference should be limited to 6 minarc for uncrossed and crossed disparity.

Jones etc al. [48] Projection Depth from Screen Limit

Experimentation provided by Jones etc al. [48] resulted in a derived limit for projected disparity of 100mm for crossed and uncrossed disparity. This limit is further discussed within Holliman [41].

Woods et al. [119] - Screen Parallax Limit

Woods et al. [119] discovered that 75% of observers could fuse images with a limit on screen parallax of uncrossed: 40mm and crossed: 45mm.

Lipton [60] - Vergence Difference Limit

Lipton [60] suggests that in order to reduce visual fatigue the limit proposed by Valyus [98] of $1.6^{o}(96minarc)$ for vergence difference should be reduced threefold to give a limit of $32minarc$ for both uncrossed and crossed disparity.

3.4.4 Disparity Variation

Discussed within Harris et al. [34][35] is the efficiency of binocular disparity achieved through experimentation using *random dot stereograms*. Random dot stereograms [57][34] consist of a series of dots in a both images which are randomly placed with a level of disparity between the pair. They only contain depth information in the form of binocular disparity. This means they are a useful tool in assessing the influence of binocular disparity. Experimentation using such images has demonstrated that observers with functional binocular vision are able to perceive depth in the absence of other depth cues. Hence, binocular disparity can be utilised even in the absence of monocular cues, contrast and prior knowledge of the observed scene.

Research has discovered that the efficiency of the visual system to detect binocular disparity is approximately 20% when the random dot stereograms contain $< 30dots$. When the number of dots in the random dot stereograms is increased, the efficiency of the visual system to detect binocular disparity correctly steadily decreases to $< 2\%$. This implies that the brain is unable to use all of the binocular disparity information in a presented scene. The measured efficiency is increased by disparity correlations [36] where dots of the same disparity levels are place in rows. This suggests that there is perhaps a local interaction of disparities; a summation and averaging mechanism for

areas within the viewing scene that are spatially located together. Hence, the amount of different disparities, the randomness of disparity arrangement and the general organisation of disparity in an image can affect the way the image is perceived. The human visual system in general finds it difficult to interpret randomly spread disparity levels. Such an image may prove uncomfortable viewing.

Assessing the *disparity gradients* present in an image is one method of establishing the variation of the depicted disparity. Disparity gradient (G) is defined by Howard et al. [43] as the difference in binocular disparity between two points in a stereo pair, divided by the difference between the mean direction of the images.

Objects residing on the horopter have a disparity gradient of zero. A horizontal disparity gradient of > 2 implies that it is not possible to observe the image binocularly resulting in a monocular perception. This means that disparity gradients > 2 will cause an uncomfortable stereo viewing experience and will not be fusible. A disparity gradient of < 2 implies that an observer may be able to fuse the stereo image but it can not say this for certain. A disparity gradient equal to 2 corresponds to Panum's Limiting Case which is the largest disparity gradient for which an image could be deemed fusible.

3.4.5 Intensity Level Differences

Intensity level difference refers to a variation in the intensity level of corresponding pixels and image segments. Within a stereo image the same scene is captured twice but with a level of horizontal disparity between the captures. The disparity results from a slightly different viewpoint within 3 dimensional camera space. When captured the difference in capture position can cause natural changes in the intensity levels between corresponding points in the stereo image pair. This phenomenon would be expected in normal viewing experiences due to natural changes in the lighting conditions for each perspective

and also occluded areas [28]. However, in general the intensity levels at corresponding image points in the stereo pair should be approximately the same.

Discussions with experts in the area have revealed that humans can usually recognise differences in intensity level present when the difference consists of a few pixels grouped together and the measured difference exceeds 5%. However, no controlled experimentation has been performed by researchers to ascertain if this value is accurate.

Howard and Rogers [43] stated that differences in intensity levels between two corresponding regions can cause a visual disturbance. A difference in intensity level over an entire image, or sections of the image can be interpreted as binocular disparity or participate in the Pulfrich Effect.

The Pulfrich Effect was discovered by Pulfrich in 1922. Pulfrich states that when the difference in intensity level difference is $> 10\%$ and either the observer or the scene is moving the Pulfrich Effect can result. The phenomenon results from a change in viewing position caused by the scene or observer moving, similar to the viewpoint difference in a normal stereo image. The difference in intensity level can bring about a sense of depth in the observed scene. This can occur even when the scene contains no binocular disparity and/or is viewed monocularly. An overview of the Pulfrich effect is given in [115][62].

Intensity differences in stereo images can also cause another phenomenon known as Binocular Rivalry which was discovered by Porta (cited within [104]) in the sixteenth century and researched extensively by Wheatstone [112] in the nineteenth century. The phenomenon of binocular rivalry was introduced in Chapter 2.

Binocular rivalry [58][43][2][3] results from a binocularly observed scene containing an area of the image (or the whole image), with a large level of intensity difference between corresponding sections. Binocular rivalry results in the observer perceiving one of the images of the pair followed successively by

the other. This results in an oscillation of the two images in the observer's perception since the observer is unable to fuse the image, or a section of the image, as a whole perception. The nature of binocular rivalry means that a visual disturbance occurs and hence the occurrence of rivalry should be avoided. The intensity level difference and the area size of the object required to cause binocular rivalry are currently unknown since most experimentation involving binocular rivalry involved test images with highly contrasted (black and white) corresponding regions. However, it is generally accepted that the human visual system is able to tolerate large levels of intensity level difference without causing rivalry [10]. In correspondence with Andrews [4] he suggested that the range of luminance could be between $13 - 58cd/m^2$ without generating rivalry.

Important observations regarding the examined psychology literature is the use of relative and non-quantitative results and analysis when examining binocular rivalrous conditions. For example, it is stated that binocular rivalry results from a perception containing a difference in contrast between the corresponding regions when this difference is large. It is also stated that fusion will still result when the difference in contrast is only slight. However, the definition gives no formal contrast levels so it is not possible to determine when each of the situations occur or what the terms 'large' and 'slight' mean.

Another important aspect of the psychological analysis of the binocular rivalry phenomenon is the fact that experiments tend to be performed on prepared experimental images with limited elements in them. It is difficult to adapt these into real world images which contain huge combinations of factors. Also this increases the complexity of the problem in the context of how factors combine in the full visual scene.

3.4.6 Frame Cancellation

Frame cancellation (also known as window violation [33]) is a phenomenon first discussed by Valyus in 1966 [98]. It is caused by the physical frame of a

display or image boundary occluding the stereo projection. It is fundamentally caused by a conflict between binocular disparity and occlusion; in such cases occlusion prevails [109][29][42]. This means that the effect caused by binocular disparity is overridden by the occlusion present. In such cases, the resulting effect is the clipping of the displayed stereo image. This is a clear violation of 'real world' experiences because there are few cases in natural viewing where such a conflict exists.

Overall, frame cancellation results in a decrease in depth perception by binocular disparity throughout the image, specifically at the image or display boundary edge [110][109]. Ware [109] also states that the effect of frame cancellation is typically accompanied by a double image of the object in front of the screen; hence, causing a greater visual disturbance. From this it can be concluded that as a baseline an image is of greater viewing quality if no frame cancellation is present. Hence, it is important that an image has minimal, if any, frame cancellation.

A diagram of frame cancellation occurring at the boundaries of a display is included by both Wartell [110] and Ware [109]. Ware demonstrates a diagram similar to the one shown in Figure 3.4 depicting a circular object projected inside the frame cancellation zone. D is the distance of the observer to the screen plane, i the observer's interocular distance and w the width of the observed display. If we take Sd to represent the distance from the display plane to the point of vergence it is possible to derive the inequalities for this model of frame cancellation in Table 3.1. An explanation of the derivation of these inequalities is given in Appendix D.

The frame cancellation area is represented in Figure 3.4 as shaded zones. These inequalities demonstrate the regions of monocular viewing - where the image is only observed with one eye and hence is no longer observed binocularly.

$$Sd > \frac{2Dx}{w+i} \wedge Sd < \frac{2Dx}{w-i}$$

$$Sd > \frac{-2Dx+D(w-i)}{w+i} + D \wedge Sd < \frac{-2Dx+D(w+i)}{w-i} + D$$

Table 3.1: Frame Cancellation Inequalities

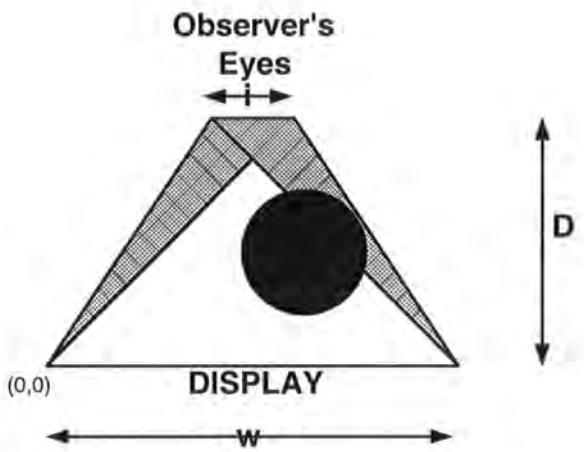


Figure 3.4: Frame Cancellation

$$Sd > \frac{2Dx-2Da}{w+i-2a} \wedge Sd > \frac{2Dx-2Da}{w-i-2a} \wedge V < \frac{2Dx}{w+i}$$

$$Sd > \frac{-2Dx+D(w-i)}{w+i-2a} + D \wedge Sd > \frac{-2Dx+D(w+i)}{w-i-2a} + D \wedge Sd < \frac{-2Dx+D(w-i)}{w+i} + D$$

Table 3.2: Frame Cancellation Inequalities

Frame Cancellation - A Simple Model

Holliman [42] describes frame cancellation as a well known phenomenon. He states that frame cancellation becomes a problem when an object exhibiting crossed disparity approaches the edge of a display, or image. He states that this problem occurs 0-20 pixels from the edge of the display, or image boundary. This is consistent with both Ware and Wartell who stated that images approaching the boundaries of the display, or image, would be cropped, hence

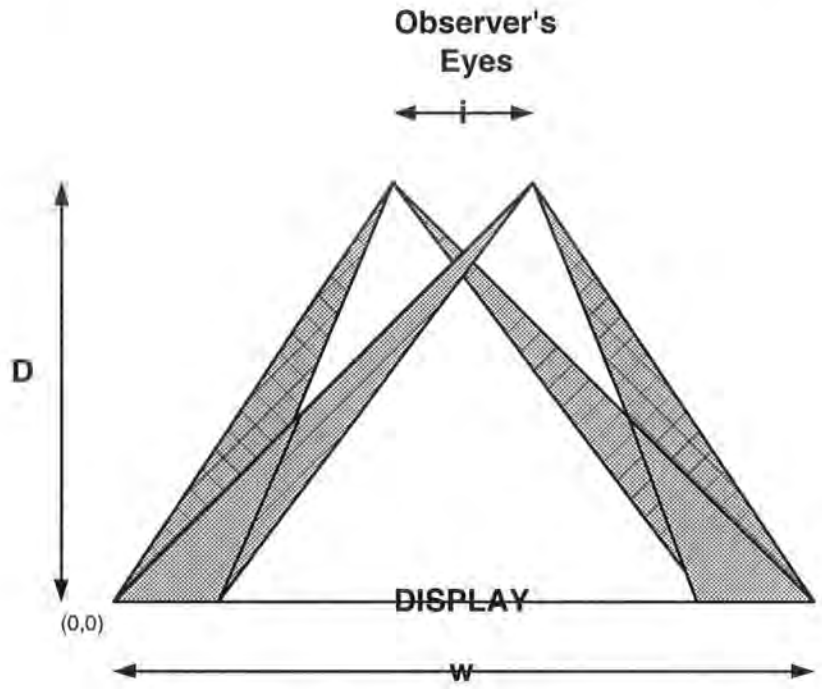


Figure 3.5: Frame Cancellation - A Simple Model of Detection

bringing about a reduction in the stereo effect. The suggested frame cancellation model is depicted in Figure 3.5. The plain shaded region depicts the detection area of the described model. D is the distance of the observer from the display plane, i the observer's interocular distance and w the display width. These regions are denoted by the inequalities given in Table 3.2 and were calculated in a similar manner to those given in Table 3.1 and explained in Appendix D. Objects projected into this zone will be observed approaching the screen or image boundary. The projected object may naturally approach and enter the monocular zone. This will result in a reduction in the stereoscopic effect caused by stereopsis for the approaching object.

3.4.7 Miscellaneous Factors

Differences in the size of the individual images and file formats (centred around the implemented compression algorithms) can cause a visual disturbance when

observed in stereo. Differences in file formats can affect the resolution of the images which is know to cause fusion and viewing comfort problems. The literature advises [65] that the size and image format of the individual images of a stereo pair are consistent and that the quality of the images is reasonable and comparable.

3.4.8 A Summary of Quality

Image quality assessment is a large research area within computer vision. The quality of stereo images is dependant upon both the quality of the single images of the stereo pair as well as the pair as an entity. Heuristics exist that attempt to determine quality of stereo images; however, these heuristics mostly operate on geometric principles images and give little consideration to human perception factors or the individual image quality. The LRC heuristic although geometrically comprehensive does not factor in analysis of depth cues, contrast conflicts, orientation conflicts or the quality factors identified by Pastoor effecting stereo images projected on an autostereoscopic display.

Many algorithms/heuristics for correspondence, disparity detection, occlusion, brightness change, contrast difference etc. have been extensively researched. These problems often have valid solutions available that will operate in a variety of image circumstances. However, many of such algorithms/heuristics are the product of industrial patents. This means that it is difficult to obtain freely available algorithms/heuristics which work as intended. The discovery, manipulation and evaluation of necessary algorithms/heuristics will provide one of the main challenges of the implementation produced as part of this thesis.

Chapter 4

A Stereo Quality Assessment Model

This chapter will develop a model for the detection of quality in stereo images. It introduces the motivation for developing the model and discusses the choices made regarding the incorporated processes. The chapter concludes with an analysis of the quality factors and an overview of the anticipated output.

The model produced as a result of this chapter will be used as a framework for a system implementation able to examine a set of stereo images and provide an analysis of their quality. The implementation of the model should provide an analysis system for assessing the quality of stereo images. It should highlight problems in the images which may cause viewing problems, specifically those that may cause the observer adverse side effects. Effectively, this could be used to analyse stereo images to assist users in determining if an image is suitable for viewing. This additionally would give stereo display developers a system users could utilise to establish suitable images to display on their systems.

4.1 An Introduction to a Model of Stereo Image Quality

The history of hardware for depicting virtual stereo is summarised in Chapter 2. In general, the future of stereo displays seems focused on providing autostereoscopic displays. The main stereo display hardware providers (Sharp, DTI, and Philips) have all invested highly in this technology. However, the current commercial successes of autostereoscopic displays is limited to research, medical and CAD applications but it is anticipated that as hardware components become cheaper and faster, the cost of the displays to consumers will reduce. This will make the displays more economically viable for everyday users and for desktop application software and games software companies to introduce systems able to utilise their features. Due to the anticipated success of these display systems the developed model will be provided on the assumed use of autostereoscopic display systems.

The growth of the Internet has resulted in an abundance of stereo images; however, most have been captured using uncontrolled procedures and have questionable quality. The introduction of autostereoscopic displays has resulted in a growing necessity for good quality stereo images. The proposed stereo image quality model is an attempt to develop an assessment procedure to identify and measure key quality attributes in stereo images and report on any possible problems detected. An implementation of such a model may be useful in assessing the suitable of stereo images for human viewing.

Regular occurrences in both psychology and computer vision research literature are references to stereo images of 'high quality'. It is difficult to determine what this phrase quantifiably means and what would be required from a system which measures it. Methods for producing stereo images in a controlled process are detailed in Chapter 3; however, there are no perfect answers for consistently achieving high quality stereo images. Indeed, there is no definition of what the phrase 'high quality' means in respect to stereo images.

4.1.1 An Introduction to Stereo Quality

An in-depth survey of the major issues affecting stereo image viewing experiences is given in Chapters 2 and 3. These issues affect the observer's perception of the quality of an observed stereo pair. Summarising from Chapters 2 and 3, the identified factors effecting stereo image quality are as follows:

- **Image size differences**
  Observing stereo images where the individual images of the pair are different physical sizes is known to cause visual disturbances. If the difference exceeds a noticeable level adverse side effects or the inability to fuse the image can result.

- **Quality differences between images**
  Differences in the quality of the individual images of a stereo image is known to cause visual disturbances and reduce the overall stereo perception quality. This individual image's quality difference can result from a difference in the image formats, capture conditions or compression rates.

- **Mismatched pairs**
  This thesis uses the term 'mismatched pair' to refer to a stereo pair where the left and right image contain no valid correspondence. Observing a pair of mismatched images will result in it being impossible to fuse the images or for a random uncontrolled and unpredictable fusion to occur. This could cause great visual disturbances and very unpredictable outcomes.

- **Excess disparity**
  Excess levels of disparity are known to cause visual disturbances when observing stereo images on stereo display systems. Stereo images containing large levels of disparity can result in fusion difficulties or adverse

side effects. Additionally, stereo pairs containing minimal levels of disparity (microstereopsis) may not be perceived to be in stereo because the level is below the threshold for human stereo recognition.

- **Excessive disparity randomness**
  The randomness of the encoded disparity present in a stereo pair can effect the overall viewing experience. Large and frequent changes in disparity throughout an image (especially thin areas of a few pixels) can cause difficulty in image fusion and can cause adverse viewing side effects.

- **Intensity level differences**
  The visual system is able to detect intensity level differences that exceed 5%. It is possible that intensity level differences can cause a visual disturbance and phenomenon such as the Pulrich effect and binocular rivalry may result. This can affect the ability of an observer to fuse a stereo image as a single perception or for random image perceptions to occur.

- **Frame cancellation**
  Frame cancellation can result when a stereo image is observed in the presence of a surrounding border and where crossed disparity is present at the left and/or right sides of image. Frame cancellation can reduce the effectiveness of the binocular disparity within an image and can result in an overall 'flattened' stereo appearance of the image.

The effect of combining these quality factors in stereo images remains unknown. For example, the resulting perception of a stereo image with both intensity level differences and disparity violations is not known. The proposed stereo quality model is built on the assumption that the presence of such factors in combination within a stereo pair will be detrimental to the stereo viewing experience as a whole and that each factor will contribute some degree

of quality degrading. Quantification of the contribution of these factors when exhibited in a single image is unknown. This means that it is not possible to give the factors' weightings when identified in combination. Hence, the model will only report on the detection of particular factors and not comment on the factors overall effect in respect to the other detected elements. Effectively, each individual quality factor will be examined separately.

4.2 A Model of Stereo Image Quality Assessment

The aim of developing a model to assess quality in stereo images is to provide an initial framework for a system that may be used by users to assess the quality of a stereo image before viewing. This type of system would be of interest to users, stereo display manufacturers and stereo image generators alike. If the system is to be effective in this purpose it should strive to fulfil the following conditions:

- It should be efficient in its ability to analyse a stereo image and produce a result in a timely manner. A system that does not produce a result in a reasonable length of time would not encourage users to use it.

- It should allow the evaluation of a large set of stereo images. If restrictions on the type of images are imposed this could mean that the predominant set of images requiring analysis are excluded - e.g. images downloaded from the Internet.

- The model should be developed with ease of installation in mind and should attempt to execute on the most commonly used operating system platforms.

- The product of the model should clearly demonstrate if a stereo image is suitable for human viewing.

These expectations of a stereo analysis system may not be possible in the developed model but should be considered and the model adapted to fulfil the criteria where possible.

The production of a model able to detect factors effecting the quality of stereo images requires decisions regarding which quality factors should be incorporated. This decision process should take into consideration the following factors:

- The factor's influence in the viewing quality of stereo images (if known).
- The ease of heuristics/algorithms generation - the thesis is bound in terms of time constraints.
- The ease of implementation - the thesis is bound in terms of time constraints.
- The speed of the implementation execution - provision of a useful stereo quality analysis tool will require that the system execution occurs in a timely manner acceptable to users.
- The availability of valid test cases - there are a small finite number of available 'high quality' stereo images.

4.2.1 The Model

The proposed stereo quality model will incorporate the following key quality attributes specific to stereo images:

- The detection of excess encoded disparity levels.
- The detection of intensity level differences.
- The detection of the presence of frame cancellation.

Detecting Excess Disparity

Excess disparity in stereo images can cause confusion for observers. In extreme cases this can result in the inability to fuse the stereo image. This results in the viewing of the stereo image becoming unattainable and/or resulting adverse side effects such as nausea, dizziness, disorientation, headaches and eyestrain etc. These problems are especially apparent for the viewing of stereo images by children under the age of 8 years. Children younger than this age can cause permanent damage to their binocular vision if they are exposed to images containing too much disparity or other image anomalies. Large disparities are a particular problem with younger observers (those under the age of 18 - when growth in humans has normally ceased) since their interocular distance is less than that of an average adult this results in a larger level of disparity being perceived. For example, the average adult has an interocular distance of approximately 65mm. If we take a screen to observer distance of 700mm and an arbitrary screen parallax of 10mm this results in a disparity level of Equation 4.1 for an adult observer. If we take a child with an interocular distance of 45mm then this results in a disparity level of Equation 4.2.

$$Sd = D + \frac{iD}{s-i} \implies Sd = 0.7 + \frac{(0.065 * 0.7)}{(0.01 - 0.065)} \implies Sd = -0.127m \quad (4.1)$$

$$Sd = D + \frac{iD}{s-i} \implies Sd = 0.7 + \frac{(0.045 * 0.7)}{(0.01 - 0.045)} \implies Sd = -0.2m \quad (4.2)$$

The possible adverse viewing conditions resulting from the presences of excess disparity levels within stereo images consequently infers that a model of stereo image quality analysis should incorporate a detection mechanism in order to be complete.

Detecting Intensity Level Differences

Differences in the levels of intensity between corresponding stereo image areas can cause visual disturbances. The extent of the visual disturbance is dependant on the severity of the intensity level difference present. Effects can range from a mere realisation of an intensity level difference to the Pulfrich effect and even binocular rivalry. These issues are discussed in Chapter 3.

The extreme case of intensity level differences results in binocular rivalry. This condition means that an observer is unable to fuse an image or image section as a stereo perception. The perception produced is also unpredictable in its nature. Binocular rivalry causes detrimental stereo viewing experiences and hence should be avoided even within confined image areas. The possible visual disturbances experienced when differences in intensity level are present in a stereo image means that a quality model should include its detection to be complete. For this reason the proposed stereo quality model incorporates an intensity level difference detection mechanism to detect the extent of intensity level differences in stereo images and report on possible regions where the Pulrich effect and binocular rivalry may be encountered.

Detecting Frame Cancellation

Discussed in Chapter 2 are the influences of depth cues present in 2 dimensional and 3 dimensional images. A significant conflict occurs when there is a discrepancy between the occlusion and binocular disparity cues depicted in an area of the image. Experiments have shown that in such cases the perception provided by occlusion will prevail; that is, the cue provided by binocular disparity will be removed from the scene perception.

Frame cancellation occurs when there is a conflict between occlusion and binocular disparity due to the presence of a frame boundary around a stereo display or stereo image. Frame cancellation occurs on an autostereoscopic display when the display projects a virtual illusion of crossed disparity whilst the

frame boundary appears behind the projected area and cancels out the depth perception given by binocular disparity. This reduces the binocular disparity effect in the area concerned and additionally causes a flattening appearance in any surrounding crossed disparity regions. This means that frame cancellation reduces the influence of binocular disparity in stereo images and hence a method for its detection would be beneficial to a system able to detect quality within stereo images.

4.2.2 An Introduction to the Model

Figure 4.1 is the proposed process model for the assessment of stereo image quality. The model incorporates the detection of the above discussed factors and formulates the following system processes:

- **Disparity Map Generation** (optional) - Required if the generation of a disparity map is necessary (in the case where no groundtruth or no other generated disparity map is present). The rectified image pair must be in PPM format. This functionality is provided by the use of third party software StereoMatch [91].
- **Disparity Limit Analysis** - Analysis is performed using a rectified stereo pair and corresponding disparity map. These are used to calculate the amount of disparity at each pixel location in the image. These disparity levels are then compared to various human factor limits to measure percentages of violation.
- **Intensity Level Difference Analysis** - Analysis is performed using a rectified stereo pair and disparity map to determine the intensity level differences between reported corresponding points. A percentage of difference in intensity level between the images is calculated.
- **Frame Cancellation Analysis** - Analysis is performed using a rectified stereo pair and corresponding disparity map to determine if crossed disparity is located within the frame cancellation zone.

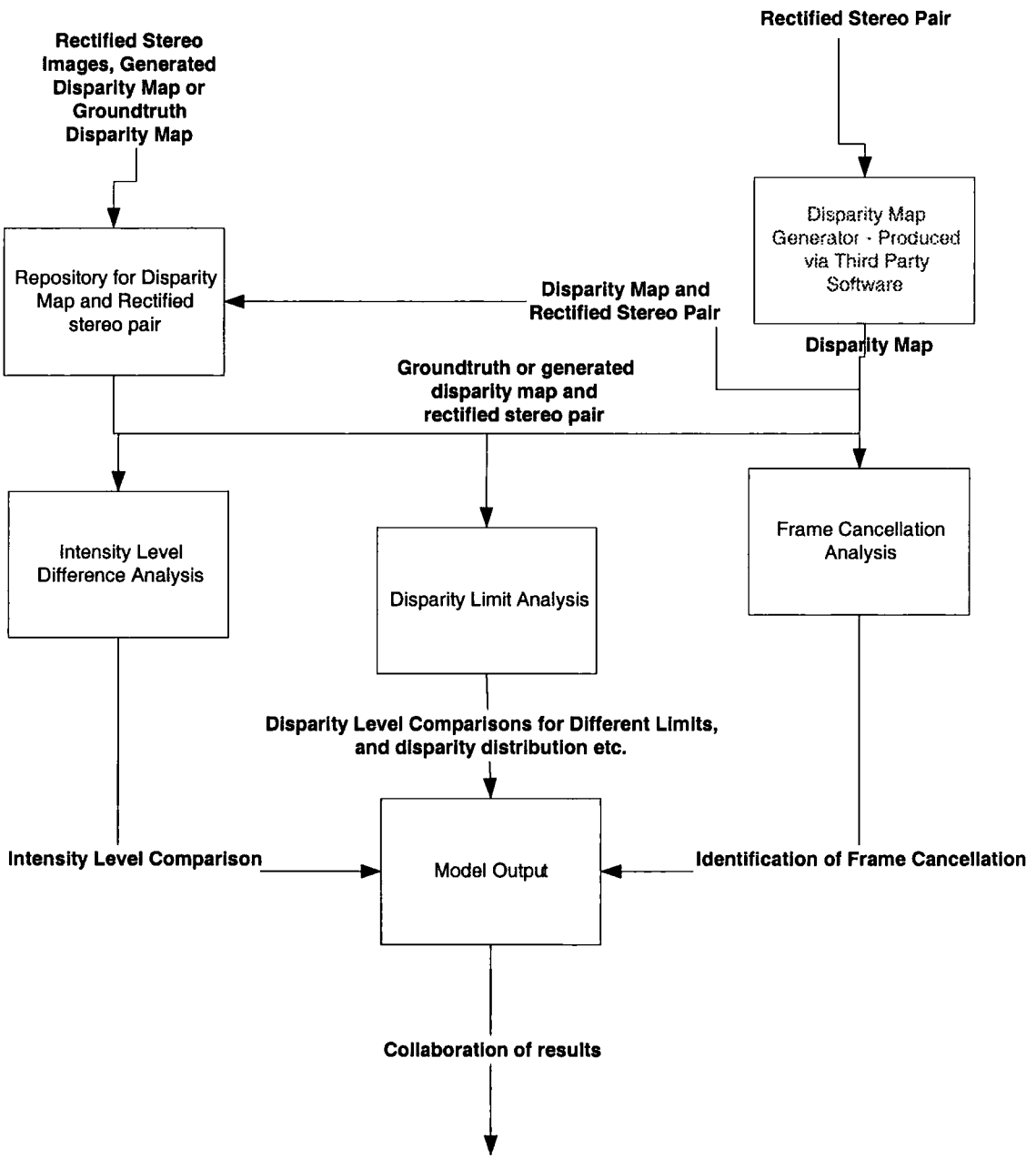


Figure 4.1: A Model of a Stereoscopic Quality Analysis System

The proposed model takes as an input a pair of rectified stereo images and where available a valid generated or groundtruth disparity map. Using these inputs it analyses the stereo pair in terms of disparity level violations, intensity level differences and the presences of frame cancellation. The output of the model consists of the original unaltered stereo image and an unaltered disparity map (if supplied to the system), any generated images as part of the implementation process and a report detailing the results obtained from the stereo image analysis.

4.3 Model Components

The model processes and the metrics formulated within them will be concisely summarised in this section. Chapter 5 will go on to further discuss the details involved in the implementation of the model.

4.3.1 Disparity Map Generation

The ability to analyse a stereo image requires knowledge of the correspondence between the individual left and right images. Stereo pair correspondence can be illustrated using a disparity map. As discussed in Chapter 3 a disparity map is a graphical representation of the correspondence present in a stereo pair. Usually a disparity map is produced based on either the left or right image of the pair. Each pixel in the image is assigned a greyscale intensity level. The intensity level is used to represent the horizontal displacement to the corresponding pixel in the matching image. This means that a disparity map can be used to calculate the implied correspondence between the left and right images of a stereo image.

The selection of a disparity map generation system is a crucial process in the development of a stereo image quality analysis system. The ability to determine a stereo image's correspondence is necessary for most stereo image

analysis procedures. In order to effectively plan the processes of the model it is essential that the resulting quality and limitations of the disparity map generation tool are known prior to model formulation.

As discussed in Chapter 3 it is possible to generate disparity maps either via groundtruth techniques or via the use of heuristics. The proposed stereo quality model aims to analyse stereo images such as those downloaded from the vast array of images available via the Internet. These images are usually captured using uncontrolled techniques. This means that it is unlikely that groundtruth images will be available with the stereo image pairs. It is not possible to generate groundtruth disparity maps post scene capture. To generate disparity maps post scene capture it is necessary to employ a disparity map generation heuristic.

Current disparity map generation systems are able to produce valid disparity map images either for a subset of image pairs or in the presense of necessary configuration requirements. The proposed model's objective is to analyse the quality of stereo images which are already in existence and where the capture method can not necessarily be established. This limits the choice of available stereo disparity map generation systems.

Several applicable systems were tested. These include StereoMatch [91], Tina [71] and Galdalf [97]. These systems purported to be able to generate disparity maps for rectified stereo image pairs. Each of these systems were available on an unsupported open source basis. Tina and Gandalf both proved difficult to install under the Linux operating system platform and included many software version compatibility requirements. Installation difficulties proved their practicality to be questionable for use in the model since it would prove difficult to incorporate them into a portable and usable system. In addition, the disparity maps they generated were of varying quality and their purported functionality was questionable.

StereoMatch like Tina and Gandalf operates under the Linux platform. It is

able to produce disparity maps from rectified stereo pairs in the presence of valid configuration information. Testing proved that such configuration data was obtainable using various image analysis techniques and that the disparity maps produced were of a good enough quality to allow some valid testing via the other procedures of the model. StereoMatch was chosen as the system to be incorporated into the model for the provision of disparity maps.

4.3.2 Excess Disparity Level Determination

Twelve limits on the level of disparity encoded in stereo images are discussed in detail in Chapter 3. Each of the limits stipulate an advised minimum or maximum limit on the level of disparity that should be present in stereo images. Exceeding the limit means that either the presences of binocular disparity becomes undetectable or the level of disparity results in diopia and/or other adverse viewing problems. The proposed model will use a selection of these limits to assess the levels of disparity in the examined stereo pairs.

The selected limits allow the formulation of specific metrics which can be used to determine a percentage of disparity detected within the images. The metrics incorporated into the model have been selected to utilise the widest range of available limits, incorporating both minimum and maximum disparity recommendations. Limits at both extremities of the 12 examined disparity recommendations have been selected along with a selection of proportioned limits within the range. Those limits used most frequently within the literature have been selected over those of a similar value. The selected limits are as follows:

1. **Siegel and Nagata** [92] stated a limit on microstereopsis. A limit of 1mm of screen parallax is stated as being the minimum detectable by humans. Disparity less than this limit will be reported by the system.

2. **Yeh and Silverstein** [121] state that the level of disparity should not exceed 27 minarc crossed and 24 minarc uncrossed of vergence difference

(for and exposure time of $< 200ms$). The system will report disparity in excess of this limit.

3. **Williams and Parrish** [116] state that the level of disparity should not exceed 25% for crossed and 60% for uncrossed disparity measured perpendicularly from the display plane. The system will report disparity in excess of this limit.

4. **Farrell and Booth** [23] state that the screen plane to point of vergence (distance perpendicular to the screen plane) must not violate $+/-$ 0.75Diopters. The system will report disparity in excess of this limit.

5. **Woods et al.** [119] discovered that 75% of observers were able to fuse images where the screen parallax was between 40mm for crossed and 45mm for uncrossed disparity. The system will report disparity in excess of this limit.

6. **Lipton** [60] suggests that in order to reduce visual fatigue a limit of $1/3rd$ should be applied to Valyus'[98] 1.6^{o} limit imposed on vergence difference. This equates to a limit on vergence difference for both uncrossed and crossed disparity of 32 minarcs. The system will report disparity in excess of this limit.

The application of the selected disparity limits to images displayed on autostereoscopic displays makes them dependant on variables introduced by both the display and the observer. Variables affecting the limits when applied to the use on autostereoscopic displays include: the display's horizontal pixel size; the optimal observer's viewing distance; the observer's interocular distance and the width of the display. It is possible to determine how these factors effect the viewing of a stereo image when examining the equations given in Chapter 2.

The stereo image analysis systems currently available provide limited support for the detection of encoded disparity levels present within stereo images. In general, the detection of excess disparity in stereo images observed on autostereoscopic displays is performed by experienced users or via an observer

suffering adverse side affects. In extreme cases the excess amount of disparity present may result in an inability of the observers to fuse the images. In such cases it is obvious to even inexperienced users that the observed images include problems.

The incorporation of human factor limits, established from psychology experimentation, into a system which is able to detect excess disparity levels is a novel application of disparity detection. A survey of the available stereo image analysis systems did not reveal any systems which combine the use of psychologically assessed human factors and a method to detect excess disparity levels within images.

Incorporation of a process based on assessing disparity levels via comparisons to human factor limits provides the model with a useful detection mechanism to prevent users observing stereo images which could cause viewing problems. This is especially important if children are to be exposed the to images. Avoiding the effects associated with the viewing of poor quality stereo images (or at least providing a mechanism to warn people of their presence) would be paramount for companies supplying autostereoscopic display systems both in terms of their reputation and the possibilities of litigation.

4.3.3 Intensity Level Analysis

Intensity level differences are present in stereo images when corresponding pixels are a different level of intensity. In colour images this can involve differences in the intensity levels present in the red, green and/or blue channels. This means that a difference in intensity level can also cause a difference in the perceived colour.

A quantitative analysis of the visual disturbance present in both greyscale and coloured images in current research is limited. The differences required and the size of the area to be recognised by the human visual system is unknown.

However, general observations and discussions with experts in the area show that most observers are able to detect intensity level differences of 5% or less. These differences can be detected by observers even when they are present in very small areas of an image (a few pixels).

Intensity level differences between corresponding pixels/sections in excess of 10% can cause problems such as the Pulfrich effect. This effect can result in an incorrect perception of depth in moving scenes or when the observer moves in relation to the scene. This additional depth may cause conflicts with the depth cues already present and could result in visual side effects or a different overall perception of the image.

At the other end of the image difference spectrum is binocular rivalry. As discussed in Chapter 2 and 3 binocular rivalry occurs when there is a large difference in intensity level between image regions. The size of these regions and the required differences in intensity level are unknown. However, research has shown that binocular rivalry causes large visual disturbances and hence a method for its detection would be well sought.

Natural Occurring Intensity Level Differences

Intensity level differences between corresponding image regions occur naturally in everyday human visual perception. These intensity level differences can be caused due to the interocular distance. This causes each eye to observe a scene from a slightly different perspective. In some scenes this can involve the left and right eye observing totally different aspects of the visual environment. The differences are known as occlusions. At occlusion areas there is no valid correspondence.

Occlusion occur naturally in stereo images and provides a beneficial cue to depth within a scene. The method by which it provides this benefit is predominantly unknown and no quantitative analysis has been performed to establish its benefit and limitations.

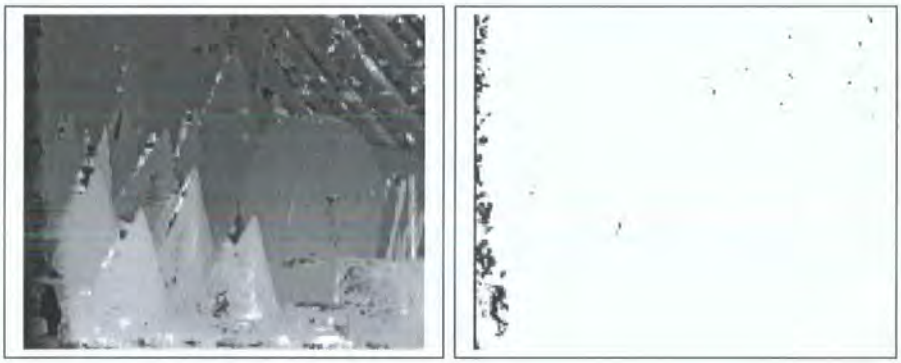


Figure 4.2: Occlusion Map

Highlighting intensity level differences in a stereo pair could highlight both the areas where intensity level problems are present and possible occlusion areas. In some disparity map generation systems occluded areas are also detected. From this analysis an occlusion map can be constructed. StereoMatch provides a method for determining occluded areas. The system flags those pixels for which it is unable to determine correspondence. This is effectively areas which it purports to be occluded. However, it will also include areas where the system failed to find the correspondence. An example of the disparity map generated by StereoMatch and the implied occlusion zones is given in Figure 4.2.

The use of StereoMatch means that many of the occlusion areas of stereo images will be eliminated from the disparity map. Instead they will be highlighted as areas where it was not possible to determine correspondence. This results in it being possible to highlight areas of intensity level differences in stereo pairs without incorporating those areas caused by occlusion. The resulting difference in intensity level will give an indication of possible problem areas in a stereo image. The model proposes to determine the difference in intensity level in a stereo image and particularly highlight intensity level differences found in the following ranges:

- $> 5\%$ - Detectable by observers.

- > 10% - In motion images or with a moving observer the Pulrich effect may result.
- > 70% - A possibility that binocular rivalry will occur.

4.3.4 Frame Cancellation Analysis

Frame cancellation is a well-known phenomenon occurring when observing stereo images whilst using a display system (such as an autostereoscopic display). The effect is also seen when observing a stereo image present with a surrounding image border.

Realistically, when viewing the world using virtual binocular disparity the phenomenon of frame cancellation will be present if cross disparity is present near the edge of the image. However, none of the examined stereo image systems provide a mechanism that allows for its detection. In general, the literature is sparse in its references to frame cancellation. Very few papers discuss the phenomenon and those that do merely suggests a suitable definition but do not given indications of its influence in observed stereo images.

Professionals producing stereo images often ensure that the left and right image boundaries contain little disparity to prevent the effects of frame cancellation. However, since the model is to be applied to previously captured images it cannot be assumed that no disparity is present in the left or right border image regions and hence a detection mechanism is necessary. Additionally, such a model of frame cancellation and its incorporation into a stereo image quality assessment model creates a novel aspect to the work.

Discussions with Holliman [42] revealed that his viewing experiences using autostereoscopic display systems and that of his colleagues suggest that the frame cancellation zone would be a typical 0-20 pixel region at both the left and right side of the display. This would be apparent throughout the total vertical display area. This is discussed in Chapter 3.

Observations have shown that in addition to problems in the frame cancellation zone a stereo image containing regions of crossed disparity at the left and right boundaries of a stereo image may gain a flattened depth appearance in the areas surrounding this region. The influence of this 'flattening' on the image is quantitatively unknown but experts in the field state that this phenomenon causes a reduction in the binocular disparity cue in the frame cancellation zone, in the regions near the frame cancellation zone and perhaps even affecting the entire image. This flattening is caused by an absence of disparity at the edges of stereo images when the brain expects disparity to be present due to the content of the image. The absence of this expected behaviour results in a reduction in the overall perception of disparity in the viewing experience, since the brain will favour the information provided by occlusion over that of binocular disparity.

The purpose of using autostereoscopic displays as opposed to normal 2 dimensional displays is to provide the additional benefit of binocular disparity cues. This is in addition to the 2 dimensional cues normally present in images. The addition of binocularly disparity gives a more complete scene with a range of different depth perception cues. Theoretically, this should give a richer viewing experience. However, a reduction in the binocular disparity cue as a result of using autostereoscopic displays could cancel out the purpose of using the displays. Hence, the presence of frame cancellation should be avoided where possible.

4.4 Summary

4.4.1 System Processes

The incorporation of human factor limits, into a process able to detect the presence of excess or minimum disparity levels, is a novel method for determining disparity levels in stereo images. Other available stereo analysis systems provide limited detection to report on the presence of possible non-fusible or

non-detectable disparity levels. Those that are available are mostly based on the inability of the system to provide valid correspondence.

The detection of intensity level differences in stereo images is provided by some stereo image analysis tools. The detection of intensity level differences by these systems is usually used to facilitate in the detection of other image anomalies. Examination of stereo analysis systems has not determined a system that evaluates the presence of intensity level difference in stereo images in order to assess their quality by highlighting the percentage of difference and possible occurrences of the Pulfrich effect or binocular rivalry.

The incorporation of a detection process for frame cancellation is an attempt at providing a quantitative definition of frame cancellation. Incorporating this into a detection mechanism for frame cancellation provides the system with a novel procedure. None of the tested stereo analysis systems available provide this functionality and no other definition for the influence of frame cancellation has been found in the literature.

4.4.2 The Quality Model

The purpose of developing a stereo image quality model is to provide a framework for a system implementation that is able to detect possible viewing problems in stereo image pairs. This could improve the overall viewing experience when observing images on autostereoscopic display systems by alerting users to problem images. Highlighting problems in stereo images which could cause the stereo effect to be reduced or the observer discomfort would be specifically beneficial to companies supplying stereo devices, companies developing software to generate stereo images or stereo image generation companies.

A system able to assess the quality of stereo images will be important if the popularity of autostereoscopic display systems grows in the future. In such cases the system could be used to provide a basic framework for producing a

more comprehensive system to assess the quality of stereo images. This may aid in the prevention of conditions associated with the viewing of poor quality stereo images. This includes both temporary and permanent side-effects associated with their use.

Chapter 5

Implementing a Stereo Image Quality Assessment System

This chapter examines the stereo quality model developed in Chapter 4, specifically addressing the implementation issues associated with producing a software system to analyse quality in stereo images. The chapter begins with an overview of implementing the software system. It then examines the individual implementation issues of the main processes of the model: disparity map generation, disparity limit analysis, intensity level differences and frame cancellation. The chapter concludes with an overview of the image quality system and a discussion of the anticipated results format.

At the end of this chapter a system implementation diagram will be produced, including the technical processes and additional features of the system.

5.1 An Introduction to The System

The system takes a stereo image to be analysed as an input. The individual images of the stereo images permitted for analysis in the system are restricted in their image formats. The formats permitted by the system include Portable PixMap (PPM), Joint Photographic Experts Group (JPG) and Portable Network Graphics (PNG) for colour images with the addition of Portable Greymap (PGM) for greyscale images. These image formats are constructed in layers where intensity levels are represented by a value in the range 0-255.

Simplistically, colour images are encoded in a three dimensional matrix where the third dimension has three elements, each corresponding to the Red, Green and Blue (RGB) image channels. The first dimension corresponds to the width of the image and the second the image height. In the matrix, an entry is made in each of the RGB elements. These are assigned a value in the range 0-255 corresponding to the intensity level of the pixel for the respective colour. In PGMs this encoding does not involve the use of colour and hence the third dimension of the matrix is limited to only one unit in size.

JPGs and PNGs differ from the other discussed image formats due to the implementation of compression. JPG compression results in a slight alteration to the data depicted in the original uncompressed image. This compression enables the image format to achieve images of smaller file size but to remain similar in terms of content to the uncompressed state. PNGs also use compression to reduce the overall file size but the image content remains unchanged from the original images. The other discussed image formats result in the image remaining in raw uncompressed formats. This means that the images require more disk space for storage but remaining intact and unchanged from the originals. Surveys of the image formats of PPMs, JPGs, PNGs and PGMs can be found in [84][7] and Table 5.1 summarises this information.

The system development is produced primarily in MATLAB. MATLAB is a programming language specifically designed for manipulating matrices and

Image Format	Colour Encoding	Image Compression	Image Changes
PPM	Yes	No	No
PNG	Yes	Yes	No
JPG	Yes	Yes	Yes
PGM	No	No	No

Table 5.1: Image Formats

performing mathematical calculations; hence, it is well suited for image processing analysis. MATLAB is able execute under both the Windows and Unix operating systems. Although MATLAB is not the most efficient runtime programming language, it is very adapted to image analysis. It has the built-in ability to easily read, write, store, iterate and manipulate images. Hence, MATLAB was determined to be a suitable choice for the implementation of the stereo quality model.

5.2 Initial System Checks and Assumptions

Discussed in Chapter 3 are the visual disturbances that can result from viewing stereo images containing errors. A subsection of these error factors have been incorporated into the system to provide an initial process in the stereo quality analysis. The following checks will be made:

- **Image sizes** - Differences in the individual image sizes can cause visual disturbances to an observer and hence should be avoided or kept to a minimum. The system provides an initial check to ensure that the left and right image are of similar size, within 5% for both the width and height of the image. Any stereo image with individual images exceeding this limit will be rejected by the system and the stereo quality analysis process will cease.

- **Compression of individual images** - Differences in the image format of the individual images can cause differences in the resulting quality of the stereo images. This may be due to the application of compression to some image formats. This should be avoided to prevent visual disturbances caused by differences in the image's correspondence. Correspondence differences can be introduced due to compression summation in the images. The system provides an initial check to ensure that the individual images are of the same format and rejects images with differences.

- **Size of the stereo image** - The system imposes limits on the size the of images it is able to analyse. Tests performed using stereo images have discovered that the system is efficient for the analysis of images up to 900 * 750 pixels in size but images exceeding this size in appropriate situations can cause a large degradation in system performance. This is due to the speed of execution and CPU and memory usage. Stereo images in excess of this size will be rejected by the system.

- **Corresponding images** - Current research results in it not being possible to determine if a stereo image matches. This matching problem is easily solved with human interaction. The human brain is able to simply determine if two images are of the same scene with some level of disparity between them. However, this process is a difficult problem in computer vision. This thesis will not attempt to address this problem and makes the assumption that the submitted stereo pairs contain some correspondence.

- **Rectified Images** - The analysis of stereo images using the model described in Chapter 4 requires that the stereo images be rectified. In non-rectified images, it is possible that the individual images are not aligned correctly. This means that it may not be possible to generate valid disparity maps or achieve accurate correspondence. The implemen-

tation assumes that the images are rectified and does not check to ensure that this is the case.

5.3 Disparity Map Generation using StereoMatch

The process of generating disparity maps is a comprehensively researched area. As discussed in Chapter 4 the disparity map generator software utilised by the system is StereoMatch [91]. StereoMatch runs as a series of scripts with the ability to generate a disparity map from a stereo pair.

The generated disparity map from StereoMatch is in PGM format. PGM format gives a range of intensity level of 0-255 units. In StereoMatch a value of 0 is used to depict a pixel where it was not possible to determine the pixel correspondence. This may result due to a occlusion within the stereo pair or areas where the system is unable to determine a valid match.

If we treat each individual image of a stereo pair and disparity map as a 2 dimensional grid with the coordinates (x, y); that is, the horizontal pixel location is assigned an x value and the vertical the y value. StereoMatch generates a disparity map based on the left image of the analysed pair. This means that the disparity map and left image of the pair have the same corresponding pixel coordinates. Hence, a pixel located in the left image at (x', y') will have a pixel in the disparity map at (x', y') corresponding to the disparity information for that pixel. This means that the disparity map can be used to formulate correspondence between the pair. This is demonstrated in Figure 5.1.

Equation 5.1 demonstrates how to calculate the disparity implied from a disparity map; where IntensityLevel(x, y) is the intensity level detected at point (x, y) in the disparity map and ScaleFactor is the scale factor applied to the disparity map. Since it is assumed the images are rectified the difference in

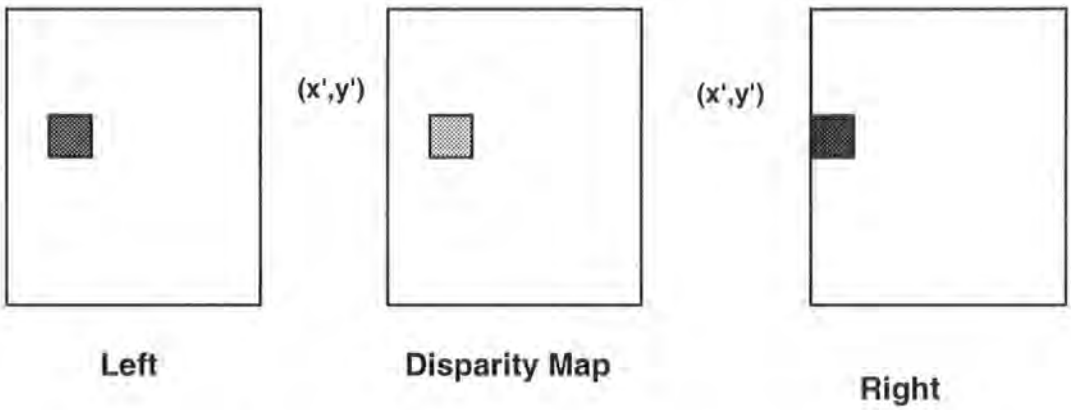


Figure 5.1: Disparity Map Correspondence

the y coordinates will be zero and hence the value of the x component of the right image can be calculated as given in Equation 5.2.

A value of $x_r < 1$ implies that there is no match present and that the image will be viewed monocularly in the left eye only.

$$\begin{bmatrix} x_r \\ y_r \end{bmatrix} = \begin{bmatrix} x_d \\ y_d \end{bmatrix} - \begin{bmatrix} \text{IntensityLevel}(x_d, y_d) * \text{ScaleFactor} \\ 0 \end{bmatrix} \tag{5.1}$$

$$x_r = x_d - \text{IntensityLevel}(x_d, y_d) * \text{ScaleFactor} \tag{5.2}$$

The scale factor determines how a change in intensity level of the disparity map is interpreted. If a scale factor of $\frac{1}{8}$ is used this represents that an intensity level of 8 units in the disparity map will equate to 1 pixel of horizontal disparity. The scale factor is effectively applied as a multiplier to the intensity level found in the disparity map.

The operation of StereoMatch requires the provision of two configuration files. One configuration file states the individual stereo pair images, the other the range and scale factor. The range determines the search space from the left to right image that the system will analyse. For example, a range of 0-150 will result in a search of up to 150 pixels to the left of the pixel coordinate in

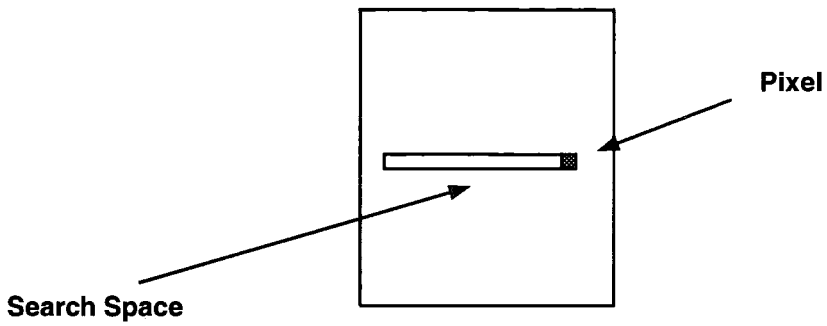


Figure 5.2: Search Space

the right image. Figure 5.2 demonstrates this searching behaviour where the shaded pixel represents the selected pixel and the resulting search space in the right image.

The required stereo image calibration information can be successfully calculated from camera calibration information associated with the stereo pair's original capture. Techniques for determining these variables from camera calibration are discussed in Montgomery et al. [67]. In the case where no camera calibration data is present, these attributes must be determined via image inspection and the generation of multiple disparity maps. In such situations, various limits and scale factors should be tested. The resulting disparity maps can be examined to ensure that the maximum detailed disparity map is achieved.

A method of determining suitable configuration information has been developed through the observation of large numbers of stereo image pairs. This process involves viewing the stereo images and determining by inspection a reasonable starting range for the encoded disparity levels and a suitable scale factor. This can be used as a method to derive the initial values to be used in StereoMatch. The range and scale factor can be adapted from these starting values to achieve a disparity map that is both efficient to produce and allows for accurate correspondence determination.

5.4 An Introduction to the Implementation of Disparity Level Violation Calculation

The process of determining the encoded disparity present in a stereo image requires the presence of configuration variables. These variables include: display specific information (pixel width, display dimensions), observer specific information (interocular distance, distance from observer to display panel) and configuration information (scale factor and search space range) associated with the disparity map.

Determining the Point of Zero Disparity

An autostereoscopic display has both the ability to provide crossed and uncrossed disparity to an observer; that is, the display can project the illusion of depth both in front and behind the display panel. To achieve this for any given stereo image, a point representing zero disparity must be determined. This value will be the point at which objects will reside on the display plane. Everything less than this value will be perceived as uncrossed disparity and everything greater will be perceived as crossed disparity.

The determination of a point of zero disparity is usually hardware dependant; however, some display systems give users the ability to specify the required point. For the purpose of developing an implementation to determine disparity in stereo images, it is necessary to decide upon a method of determining a suitable point of zero disparity. This thesis will refer to the point of zero disparity as the *midpoint*. Two solutions for the determination of a midpoint have been considered:

1. **A ratio of crossed to uncrossed disparity derived from human factor limits**
   This method would utilise the ratios of crossed to uncrossed disparity depicted in each of the human factor disparity limits. For example, a screen parallax limit regarding crossed disparity of 36.4mm and uncrossed dis-

parity of 27.3mm (given by Jones [48]) would equate to a ratio of 0.57:0.43 for uncrossed to crossed screen parallax, respectively. Since the disparity maps contain ranges of intensity levels represented as a value 1-255 (0 is used to denote a pixel where disparity level determination was not possible) the ratio can be translated into a midpoint of 110. In such an example, the range 1-109 could be used to denote uncrossed disparity, 111-255 crossed disparity and 110 the midpoint.

2. **A ratio of crossed to uncrossed disparity derived from the examined image**
   This proposed method would utilise the range of depicted disparity levels present in the examined disparity map. This range would be used to define the midpoint to be the mid value between the maximum and minimum intensity values within the disparity map. For example, a disparity map with a range of 24-156 units would give a midpoint value of 66 units.

The first considered solution involves using human factor ranges. This would result in differences in the midpoint for every human factor limit. This consequently would make it difficult, if not impossible, to compare different human factors within the same stereo image. In addition to this problem, the ratio for each of the disparity limits would differ depending on the examined variable. That is, for any particular limit the ratio obtained from screen parallax, depth from screen plane and vergence difference could all differ. This behaviour is specifically apparent when considering the formulas given in Chapter 2 and 3. Hence, any implementation of a ratio implementation based on human factor limits would require a decision regarding which variable to use to formulate the overall comparison.

The second considered solution gives a fairer representation of the encoded disparity range. The calculated midpoint would be consistent throughout the analysed human factors limits since it is image dependant. In addition to this, since the midpoint remains constant for a given image it is not necessary to specify a comparison variable. This means that a fairer overall comparison

could be made between limits for each examined image. The system implementation will adopt this method in its approach for midpoint determination.

Equation 5.3 demonstrates the calculation of the number of pixels of screen disparity considering a midpoint (MP) measured in pixels. IntensityLevel(x_d, y_d) denotes the disparity map intensity level for the examined pixel at coordinates (x_d, y_d) and sp the calculated screen parallax in pixels. If the value of $sp > 0$ the examined pixel exhibits uncrossed disparity. If the value of $sp < 0$ the examined pixel exhibits crossed disparity. A value of $sp = 0$ signifies zero disparity.

$$sp = (MP - \text{IntensityLevel}(x_d, y_d)) * \text{ScaleFactor} \tag{5.3}$$

$$s = \text{round}(MP - \text{IntensityLevel}(x_d, y_d)) * \text{ScaleFactor} * p \tag{5.4}$$

The value of sp in Equation 5.3 could result in a floating point value. A display is only capable of depicting whole value pixels and hence it is only able to depict whole values of pixel disparity. This means that it is necessary to round the results. This function is depicted in Equation 5.4, where p is the pixel width and 'round' the rounding function employed.

5.5 Implementing an Analysis of Disparity Violation

As discussed in Chapter 4 the system will utilise six human factor limits to analyse the levels of disparity present in an examined stereo pair. The limits will be used to formulate metrics which will measure the level of violation for each of the limits in the examined stereo images.

The implementation examines the violation of the human factor limits for all of the pixels present in the left image with a matching pixel present in the right. Only pixels with a corresponding pixel will be considered since non-matching pixels will not participate in providing an observer with binocular disparity. Calculations will be made according to the number of pixels with a measured disparity level exceeding each of the levels denoted by the human factor limit. A percentage of the examined pixels violating the given limit is calculated for each of the limits and reported by the system.

The detection of disparity level violations involves measuring the level of disparity in a stereo image and comparing this to a particular disparity limit. This comparison is performed for each pixel in the left image of the stereo pair. Any pixel violating the disparity limit is counted and used in calculating an overall percentage of pixels with disparity levels violating an examined disparity level. In this method a pixel is merely counted as violating or not violating. This method of stating an image's violation does not give an indication of the extent of the violation present. For example, if an arbitrary disparity limit for screen parallax of 1mm were selected, a pixel with a corresponding screen parallax of 1.1mm would be counted as a violating pixel just the same as a pixel with 10mm of screen parallax. This means that the percentage of image violation would not give an overall representation of the extent of the violations present. This makes it difficult to compare images effectively. Hence, it is beneficial to determine a method that will enable effective and fair comparisons between the percentages of violation of disparity limits within stereo images.

It is essential that any devised metric for representing disparity limit violation in a given stereo image would allow a fair comparison to be made between images. The purpose of analysing disparity limit violations is to achieve an indication of the quality of viewing an image on a particular stereo display. This should allow a comparison between images regardless of their size, disparity map generation method and disparity limit adopted in the individual metric.

A solution to this problem is to use a *weighting function.* A weighting function [114] is a mathematical function that is applied to a selected section of data to change its behaviour.

5.5.1 Weighting Functions

Various weighting functions [114] can be applied to data sets to influence the data's behaviour. Different types of weighting functions are linear, polynomial, logarithmic and reciprocal.

The simplest form of a weighting function is a linear weighting function. This involves the application of a linear relationship to an examined data range. The gradient of the function is constant throughout. The result is a function that gives equal weighting throughout a data range without favouring either the small or large values in respect to a given value. The gradient of the linear function can be changed to alter the degree by which the data will change. This behaviour is demonstrated in the graph in Figure 5.3.

A logarithmic weighting function has a larger gradient for small changes in respect to a given value, x', and reduces to a smaller gradient for larger differences where it plateaus to a chosen maximum value of y. This means that as the value of x is initially changed the value of y changes rapidly. However, this change plateaus, converging onto a value of y. This behaviour is demonstrated in the graph in Figure 5.5.

A polynomial weighting function has a smaller gradient for small changes from a chosen x value and a larger gradient at larger change levels. Hence, this gives a larger weighting to larger values than those of smaller values difference. This behaviour is demonstrated in the graph in Figure 5.4.

Applying a Weighting Function to Disparity Limit Analysis

The purpose of applying a weighting function to the range of detected vio-

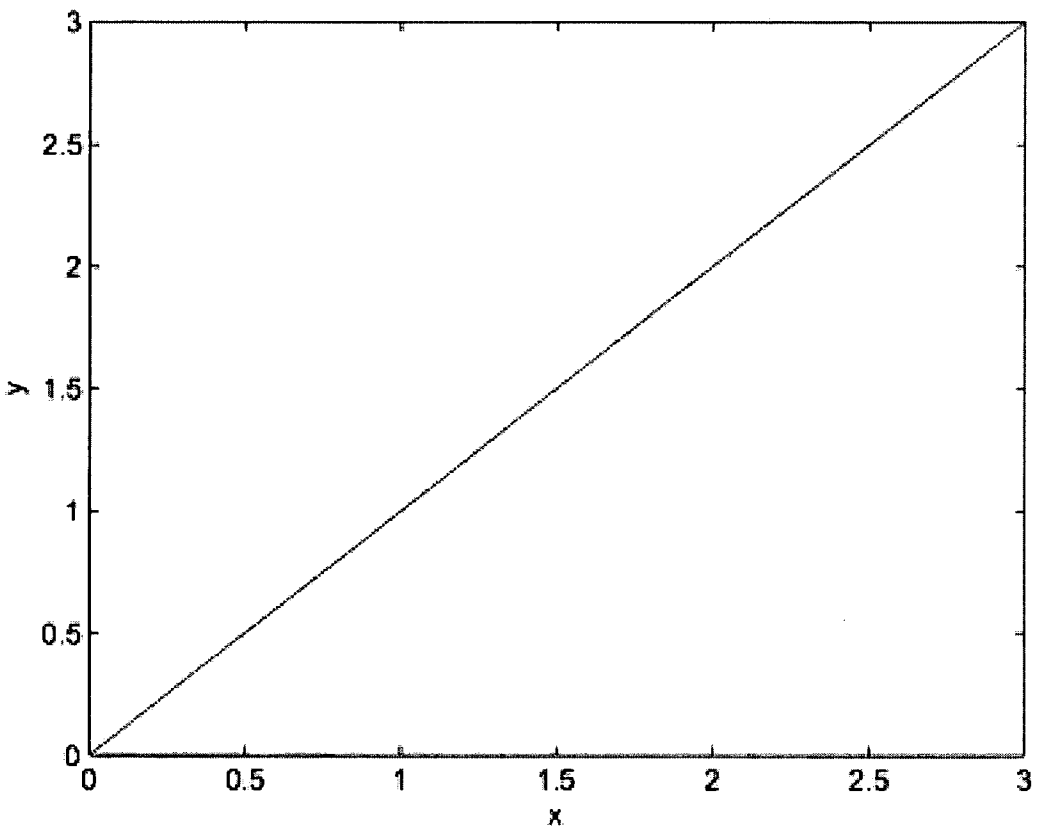


Figure 5.3: A linear graph (y = x)

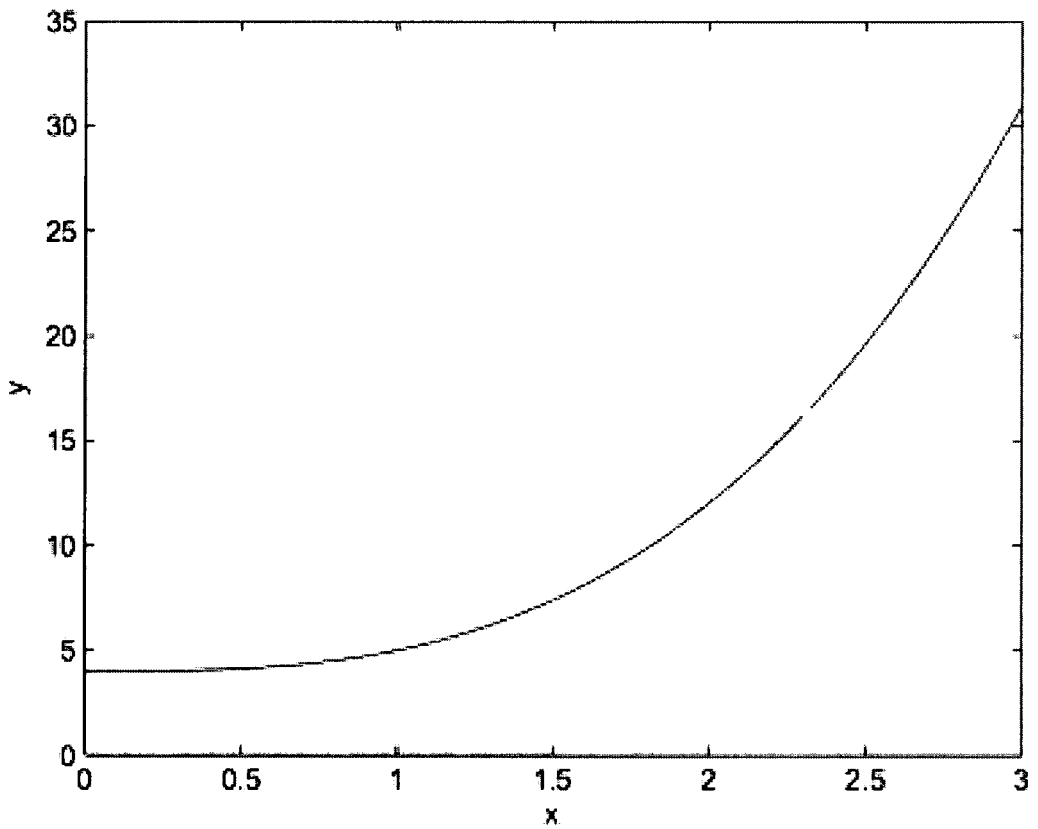


Figure 5.4: A Polynomial Graph (y = $x^3 + 4$)

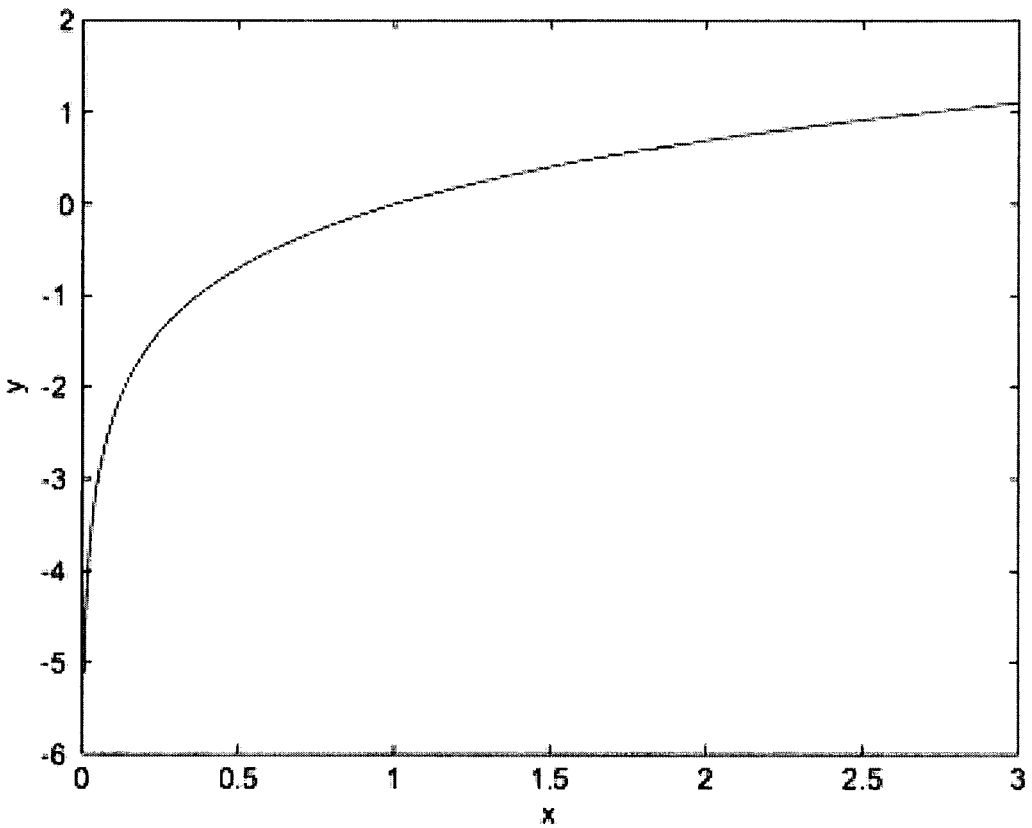


Figure 5.5: A logarithmic graph (y=log(x))

lating disparities is to give a more accurate indication of how much disparity violation is present in a stereo image pair for each of the examined limits. The larger the amount of disparity detected in comparison to a specific limit, the greater the violation that should be recorded. Hence, as opposed to a system where a non-violating pixel is assigned a value of 0 and a violating pixel 1; a scale of 0-1 could be employed so that a value of 0 denotes a non-violating pixel and a value of 1 a pixel with a large violation of the disparity limit. The value of pixels in between this range is given a value in the range 0-1 calculated using a suitable weighing function.

It would be beneficial if the devised weighting behaves so that pixels with a small violation of the limit should be penalised less than those with larger violations. This criteria is more suited to a polynomial weighting function. Hence, a polynomial weighting function could be applied to achieve a new metric for assessing the amount of encoded disparity in a stereo pair.

Various polynomial functions could be applied to the data range. These include squared, cubic and polynomial. These functions are depicted in Figure 5.6. The squared function was selected as most suitable because its behaviour means that small changes from the disparity limit could be easily weighted less than those showing greater levels of violation.

A polynomial weighting function can be normalised so that the data range is between 0-1 where 0 represents a pixel which does not violate a limit and 1 a pixel that violates a limit greater than or equal to the chosen maximum value. To implement this it is necessary to determine a suitable value of disparity which will be assigned 1 in the disparity limit violation measurement.

Choosing a Suitable Maximum Value

Four possible maximum value determination methods have been considered to achieve this:

1. **Limiting the range depending on the disparity depicted in the**

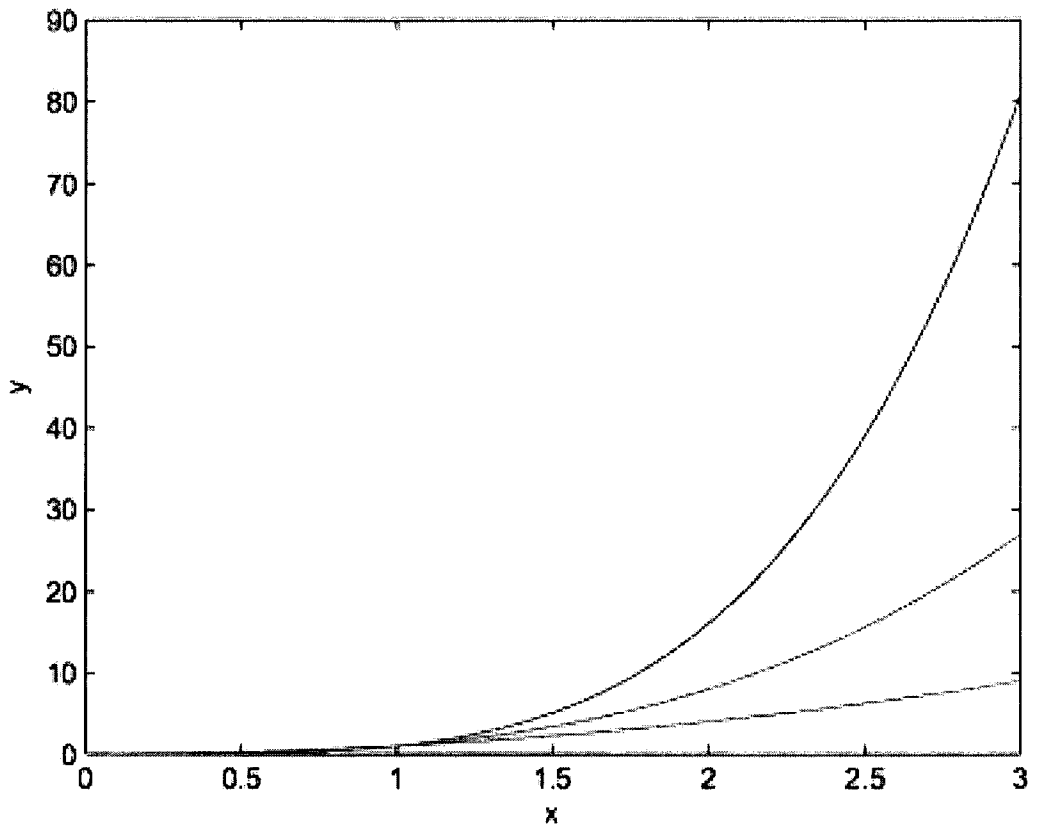


Figure 5.6: A Comparison of Polynomial Graphs (bottom - $y = x^2$, middle - $y = x^3$ and top - $y = x^4$)

disparity map

The first considered solution involves analysing the disparity map and associated scale factor to calculate the maximum possible disparity which could be depicted. For example, if a disparity map has a $\frac{1}{4}$ scale factor associated with it this results in each change in intensity level being equivalent to $\frac{1}{4}$ pixels of horizontal disparity. If we assume that the disparity map has an even distribution of disparity levels and using the decided approach for the calculation of a suitable midpoint i.e. ranging from 1-255 units the midpoint will be 128 units giving a maximum crossed and uncrossed disparity of 128 and 127 units, respectively. With a $\frac{1}{4}$ multiplier this equates to a maximum of 32 pixels for both crossed and uncrossed screen parallax. Using this assessment, the value 1 within the weighting structure would be used to denote a screen parallax of 32 pixels or more.

2. **Limiting the range depending on the display width**

   The width of the display will limit the amount of permitted screen parallax. This will impose a limit of $w-1$, where w is the width of the display in pixels. In general, stereo images will show an even distribution of disparity levels throughout the image and hence a midpoint around 128 will be selected. This would give a maximum screen parallax of half the display width (for both crossed and uncrossed disparity). That is a limit of $w/2$. However, it would be expected that very few images would contain such a large disparity encoding unless the display width was small.

3. **Limiting the maximum value depending on the image width**

   The width of an examined stereo image would also place a limit on the possible encoded disparity depicted within a given image. This would limit the image disparity to $(iw-1)$, where iw is the width of the image in pixels. In general, stereo images will show an even distribution of disparity levels throughout the image and hence a midpoint around 128 will be selected. This would limit the maximum disparity to half the width

of a given image and hence, $iw/2$. However, it would be expected that very few images would contain such a large disparity encoding unless the image width was small.

4. **Limiting the maximum value depending on the examined limit** The disparity limit can be used to calculate an appropriate maximum value by reflecting the magnitude of the violation by a multiple of the original limit. For example, if a limit of 10 units is used then a multiple of this limit could be used to denote the maximum value. In the case where double the human factor's limit is chosen this would give a value of 20 units.

The first suggested solution may cause problems in determining an appropriate unbiased maximum value. If an inappropriate disparity map is generated (one with an inappropriate scale factor and/or range) the result would be a biased stereo image analysis. For example, if a disparity map is generated with a scale factor of $\frac{1}{2}$ and a midpoint of 128 is calculated for the stereo pair. This would give a > 64 pixel screen parallax to denote a 1 unit weighting. However, if a scale factor of $\frac{1}{4}$ would more accurately represent the image this would give a screen parallax of > 32 pixels denoting a 1 unit weighting. It is possible for a $\frac{1}{2}$ to represent a $\frac{1}{4}$ scale factor disparity map (although it will include less detail if a $\frac{1}{4}$ scale factor would be more suitable) so determining the weighting scale based on the scale factor associated with the disparity map could cause errors and is not a reliable means to assess an image to allow a fair comparison.

Solutions 2 and 3 involve examining the extreme levels of disparity that could be present depending on either the image width or the display panel width. Choosing a suitable maximum that is a value of half the possible disparity could be used to formulate an appropriate maximum weighting value. However, this method would not reflect on the magnitude of the original examined limit and would give a maximum value which was constant throughout the limits. This means that in some cases it is possible that the disparity level limit is exceeded

hundred or thousands-fold before it is given the full 1 unit value. Additionally, due to the differences between each of the limits this method of maximum value determination would mean that it was no longer possible to compare stereo images of different sizes or the results from each of the limits with one another.

The fourth solution results in the weighting function reflecting the original disparity level limit and will be consistent throughout all stereo images. This means that the ability to compare limits and images is retained. Additionally, it is possible to alter the magnitude of the maximum value so that it is possible to adapt the weighting depending on the images analysed.

This method of midpoint determination has been adopted by the system. Initially, the maximum value will be chosen to be double the original limit. This must be normalised within the polynomial function to give a weighting for a non-violating pixel of 0, a pixel, p, violating between $limit <= p < 2limit$ in the range 0-1 and determined by the polynomial function, and a limit $>= 2limit$ of 1. Equation 5.5 demonstrates the squared weighting function, where the disparity limit implemented, s is the examined pixel's value using the appropriate limit variable and the resulting weight the weighted value assigned to the result.

$$weight = ((1/limit) * (s - limit))^2 \tag{5.5}$$

Use of such a weighting function means that the system is able to consider some degrees of minor limit violation and give a fairer representation of the disparity limit violations present in an examined image.

The weighting function will be applied to the Yeh and Silverstein (Y+H), Farrell and Booth (F+B), Williams and Parrish (W+P), Woods et al. (Woods) and Lipton (Lipton). Applying a weighting function to the Siegel and Nagata (S+N) limit posses a problem in the definition of the devised limit. This is due to the S+N limit being an advised minimum. For this reason a weighting

	Screen Parallax (mm)	Pixels
Y+S	5.511	18.507
F+B	34.132	114.6279
S+N	1	3.358
W+P	21.67	72.7655
Woods	40	134.336
Lipton	6.53	21.935

Table 5.2: Crossed Disparity - Screen Parallax Limits

	Screen Parallax (mm)	Pixels
Y+S	4.897	16.445
F+B	34.13	114.618
S+N	1	3.358
W+P	24.375	81.861
Woods	45	151.128
Lipton	6.529	21.926

Table 5.3: Uncrossed Disparity - Screen Parallax Limits

analysis will not be applied for the S+N limit throughout the system analysis.

5.5.2 Disparity Image Analysis on a 15" Autostereoscopic Display

The 15" autostereoscopic display described in Chapter 3 has been used to calculate the disparity map scale factors which enable each of the examined limits to become violated. The scale factors used in the images vary from $\frac{1}{8} - \frac{1}{2}$. As calculated in Table C.1 the limits translate into the values given in Table 5.2 for crossed and Table 5.3 for uncrossed disparity for screen parallax.

For a stereo image with the full range of intensity levels in the corresponding disparity map (1-255) this would result in a midpoint of 128. This gives a range of 1-127 for uncrossed disparity and 129-255 for crossed disparity with 128 being the point of zero disparity. Table 5.4 illustrates the effects of applying a scale factor to these values and the maximum amount of resulting screen

Scale Factor	Screen Parallax (mm)	Pixels
$\frac{1}{8}$	4.725	15.875
$\frac{1}{4}$	9.451	31.75
$\frac{1}{2}$	18.90	63.5
1	37.80	127

Table 5.4: Maximum Possible Disparity and Scale Factors

Scale Factor	Y+S	F+S	S+N	W+P	Woods	Lipton
$\frac{1}{8}$	No	No	Yes	No	No	No
$\frac{1}{4}$	Yes	No	Yes	No	No	Yes
$\frac{1}{2}$	Yes	No	Yes	No	No	Yes
1	Yes	Yes	Yes	Yes	No	Yes

Table 5.5: Possible Limit Violations and Scale Factors

parallax for the examined 15" display.

Comparing Tables 5.2, 5.3 and 5.4 means it is possible to determine the limit violations for each scale factor. These are summarised in Table 5.5.

Using this analysis it is possible to determine the scale factors associated with disparity maps which could result in a violation of each of the examined limits. Depending on the pair's associated scale factor for either a groundtruth or generated disparity map it is possible to determine which limit violations may result - assuming that the disparity map scale factor is appropriate for the stereo image. The larger the disparity map scale factor the greater the level of possible violations since the larger the amount of encoded disparity which can be depicted.

It is possible to use these results to reduce the calculations required on examined stereo images. For example, a stereo pair with a disparity map with an associated scale factor of $\frac{1}{4}$ would result in it not being necessary to calculate the violations of the F+S, W+P and Woods limits since violations would not be present at these levels. This saves computational costs associated with the image analysis of stereo images where violations are not possible.

5.6 Intensity Level Differences

Discussed in Chapter 4 is a model of intensity level difference. In summary, the system will detect and report on differences in pixel-to-pixel intensity difference. It will highlight specifically areas which exceed 5% (detectable by human observers), 10% (possible Pulfrich effect in moving scenes and 70% (possible binocularly rivalry). The intensity level difference will be reported as an overall percentage of image pixels violating these limits.

Mathematically, Equation 5.6 describes how to calculate the intensity level difference present at a given pixel in the left image. In this equation dx denotes the disparity at pixel coordinate (x, y) in the left image and IntensityofLeft(x, y) obtains the intensity level of this pixel. IntensityofRight(x', y') obtains the intensity level at coordinate (x', y') in the right image and abs is a function used to calculate the absolute value of its parameter.

Figure 5.7 demonstrates a greyscale stereo pair where a level of disparity is present between the two images. In the example, a pixel at coordinate $(2, 2)$ in the left image has a corresponding pixel in the right image at coordinate $(0, 2)$ - both pixels are shaded grey. There is no vertical disparity present between the two pixels but there is a horizontal disparity of 2 pixels at reference pixel $(2, 2)$ in the left image. This means that the corresponding pixel can be found at $(0, 2)$ in the right image. If the intensity level at pixel $(2, 2)$ is 125 and in the right image at $(0, 2)$ is 110 the resulting intensity level difference is calculated to be $\frac{(125-110)}{255} * 100 = 5.88\%$.

The system analyses each pixel present in the left image of the stereo pair and considers those with a corresponding pixel within the right image. It uses the sum of the differences to calculate the average intensity level difference present in the image.

A discussion of the image formats to be examined by the system is given at the start of this chapter. Examining colour images (i.e. PPM, JPG and PNG)

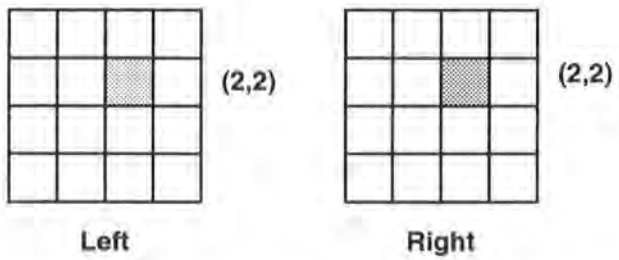


Figure 5.7: Intensity Difference - Comparison

requires an intensity level comparison for all three colour channels. An overall intensity level difference can be obtained by taking an average of the differences over all three channels. This is represented in Equation 5.7; where IntensityDiffRed, IntensityDiffGreen and IntensityDiffBlue are the intensity differences measured using Equation 5.6 for the red, green and blue channels respectively.

$$Difference(\%) = \text{abs}(\frac{\text{IntensityofLeft}(x,y) - \text{IntensityofRight}(x+dx,y)}{255})*100 \tag{5.6}$$

$$Difference(\%) = \frac{(\text{IntensityDiffRed} + \text{IntensityDiffGreen} + \text{IntensityDiffBlue})}{3} \tag{5.7}$$

For example, a pixel with an intensity level of 123, 156, 178 (RGB respectively) in the left image and a corresponding pixel in the right image with an intensity level of 23, 56, 179 would result in an overall intensity level difference of $\frac{\frac{(123-23)}{255}+\frac{(156-56)}{255}+\frac{(-178+179)}{255}}{3} * 100$ =52.42%.

5.7 System Analysis of Frame Cancellation

As discussed in Chapter 4 the model of frame cancellation to be used by the system involves detecting any crossed disparity in the left and the right side of an examined stereo image. A range of 0-20 pixels was chosen to be the detection area. Any crossed disparity found in this region would be detected and reported by the system.

The type of disparity map utilised by the system includes pixel correspondence with the left image of a stereo image. However, if the calculation of frame cancellation is only based on the left image it will be limited to detecting occurrences of frame cancellation at the right side of the stereo image only. However, in the proposed implementation the left side of the image would only be highlighted as causing frame cancellation if the crossed disparity did not exceed 1-20 pixels depending on the examined pixel. This arrangement means that frame cancellation could be prominently present in the left side of an image but not detected.

To calculate the frame cancellation present in the left side of a stereo image it is necessary to perform a preliminary step in the frame cancellation calculation process. This involves calculating the correspondence from the right to left image for the 1-20 pixels present in the left side of the right image.

To achieve right to left correspondence, initially, an analysis of the left image is performed with a pixel corresponding disparity map. The matching pixels in the right image are highlighted. During this procedure, if a pixel in the right side of the right image is highlighted and is present in the 1-20 pixel region the disparity translation is recorded. For example, if a pixel located at coordinates $(30, 10)$ is found to have a disparity level of 12 pixels from the left to right image and assuming a scale factor of 1, this would translate into the corresponding pixel being located at $(18, 10)$ in the right image. Since the x coordinate is < 20 the matrix is marked to indicate that a match is present in the 1-20 pixel region. This makes it possible to examine the highlighted pixels

to ascertain if they exhibit crossed disparity and hence are involved in frame cancellation.

5.8 Image Segmentation

It is possible stereo images can exhibit large degrees of variation in disparity level and intensity level difference throughout the image. This results in the possibility that an image contains areas where disparity level violation and/or intensity level difference are high but when the image is examined as a whole these anomalies areas are lost in the overall results. For example, in Figure 5.8 there is a small region of large disparity violation for limit present in the top left corner of the image (represented in black). However, the rest of the image contains no other violation areas (represented in white). If this image is analysed as a whole, a small percentage will be reported for the image's disparity level violation. Effectively, the violating region has been lost in the vastness of the image with the violation area being lost in the averaging process. To provide a solution to this problem, a method of simple segmentation has been proposed. To prevent large costs associated with artefact segmentation algorithms, a simple method of segmenting an image into 100*100 pixel sections has been selected. This method will be used to segment both the left image of the pair and the associated disparity map into 100*100 pixel sections starting from the top left hand corner of the image and similarly to the example given in Figure 5.9.

Using segmentation it is easier to identify regions of images where violations occur because the areas involved are smaller. Image segmentation will be used in the analysis of disparity level violations and intensity level differences so that it is possible to identify regions of interest in the examined stereo image.

Image segmentation also allows an analysis of the variance of disparity and intensity level differences throughout a stereo image. It is possible using the results for each of the image segments to analyse the disparity and intensity

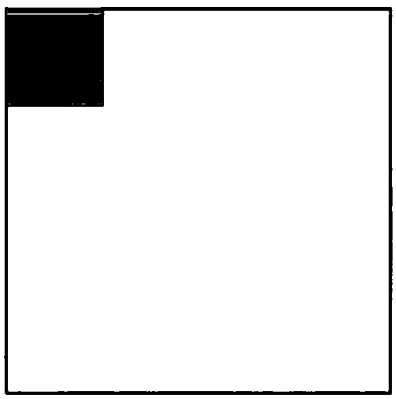


Figure 5.8: Segmentation - Locating Violations

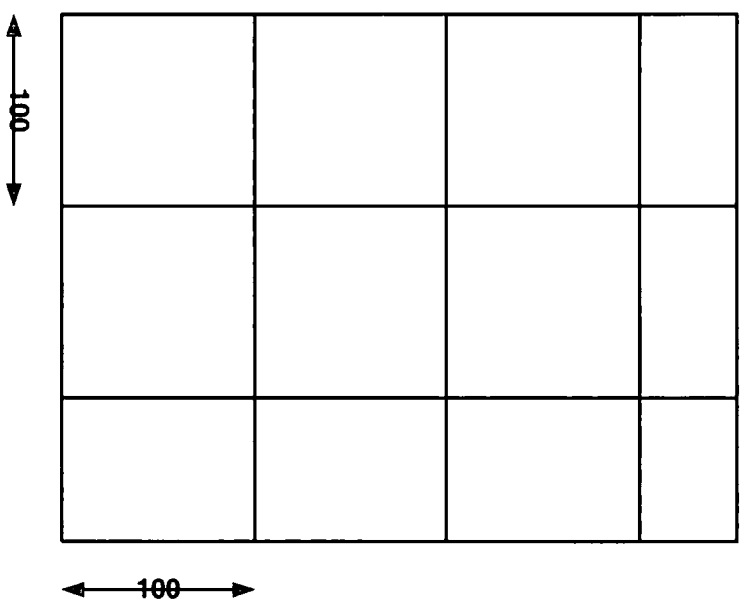


Figure 5.9: Image Segmenation

level distributions throughout a stereo image. This includes calculating the variance, range and other statistical functions on the data. This analysis will aid in ascertaining how disparity violations and intensity level differences are distributed throughout a stereo image.

5.9 Summary

5.9.1 Inputs to the System

The implementation of the model discussed in this chapter requires a number of inputs to the system before image analysis can commence. Initially, the system requires the input of a rectified stereo pair and parameters associated with the autostereoscopic display and observer. A groundtruth (or generated) disparity map where available can also be provided. This permits users to utilise groundtruth disparity maps or generated disparity maps in the system gives it greater flexibility by allowing for different disparity map generation techniques.

The parameters required by the system are based on information relating to the observer (distance from the display panel and interocular distance) and the autostereoscopic display (pixel width, observer to screen distance range, display width and display height). This information allows the system to derive an image quality analysis that is specific to the viewing situation without any prior assumptions. This makes the system versatile so that it can be used by a range of observers and autostereoscopic displays.

5.9.2 System Results

In summary, the system analysis of disparity level violation, intensity level differences and frame cancellation will be provided in a report at the end of the image analysis procedure. This will include information relating to the disparity level violations, intensity level differences and frame cancellation detected

in the examined stereo pair. The implementation of the system is summarised in Figure 5.10.

In summary, the comparisons made by the system for disparity violation level include the following:

1. Overall disparity level violation using 6 separate human factors limits. Each pixel in the left image with a corresponding pixel in the right image will be examined and the disparity level between them calculated. This disparity level will be compared to 6 metrics and used to calculate 6 disparity violation percentages for the average measurements made. These values are calculated for weighted and non-weighted measurements for the examined image. This includes separate calculations to determine if the disparity violation is caused by crossed or uncrossed disparity and a percentage of each violation given.

2. Segmentation of the left image and the corresponding disparity map into 100*100 pixel segments. The original non-segmented right image is used to provide a pixel-to-pixel comparison as per above. This allows a demonstration of the variation of disparity level violation throughout a given image.

The comparisons made within the system for intensity level differences included the following:

1. Overall pixel-to-pixel intensity level comparisons for each pixel present in the left image of a stereo pair. Each pixel's intensity level is compared to its corresponding pixel and the difference based on 256 levels. An overall percentage of intensity level violation is obtained for the stereo pair by taking an average of the values obtained.

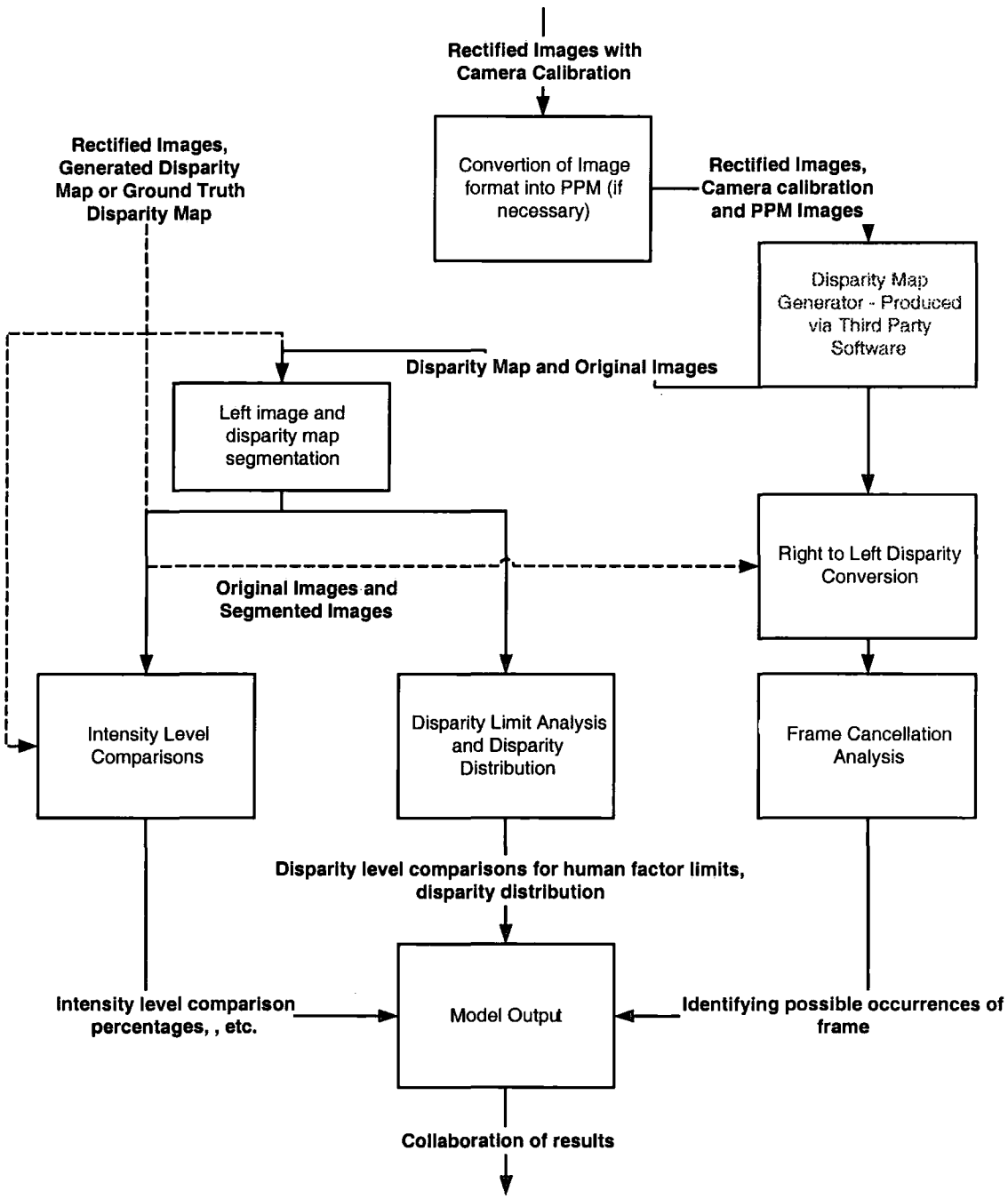


Figure 5.10: An Implementation Diagram of the Stereo Quality Analysis System

2. Segmentation of the left image and the corresponding disparity map into 100*100 pixel segments. The original non-segmented right image is used to provide a pixel-to-pixel comparison as per above. This allows a demonstration of the variation of intensity level differences throughout a given image.

The system's analysis of frame cancellation provides a measurement for an examined stereo image. The measurement is given for both the left and right side of the pair where frame cancellation can occur and consists of a percentage of the pixels involved in frame cancellation including all pixels within the bounds being examined (both crossed, zero and uncrossed disparities).

The system will output in addition to the results of the analysis, the original rectified images and the disparity map (if provided) or a generated disparity map with the relevant configuration information.

Chapter 6

Results and Image Analysis

This chapter will summarise the results and analysis produced from the examination of a set of stereo images using the developed stereo image quality assessment system. A selection of test stereo images are inputted into the system and the results from disparity limit violations, intensity differences and the presence of frame cancellation are analysed and evaluated.

An assessment is made of the error factors introduced by the use of Stereo-Match as a disparity map generator, the effect of image compression and natural variations of disparity violations and intensity level differences detected by the system.

6.1 The Test Images

The images selected for analysis within the devised stereo quality assessment system are 14 stereo images that highlight the diverse range of stereo images currently available. The selected images are all camera captured and vary in the quality of capture method.

The availability of stereo pairs with corresponding groundtruth is scarce and hence the Middlebury images [90] have been extensively utilised throughout the system evaluation. In addition to the Middlebury images, stereo images produced from Sharp [70], a scanned image and an image downloaded from the Internet have also been utilised.

Six Middlebury images (Barn1, Barn2, Bull, Poster, Sawtooth and Venus) were selected from the small image range (approximately 430 * 350 pixels). These images are rectified and have a groundtruth disparity map associated with them with a scale factor of $\frac{1}{8}$.

Two sets of large Middlebury (SmallCones, Cones, SmallTeddy and Teddy) stereo images (450 * 375 and 900 * 750 pixels) were also selected. These images are rectified and have a groundtruth disparity map associated with them with a scale factor of $\frac{1}{2}$ and $\frac{1}{4}$. The 1800*1500 pixels stereo images which are also part of the large Middlebury stereo image suite were not analysed due to their size being restrictive in disparity map generation with StereoMatch.

Two images (Sharp Farm and Sharp Hand) were selected from those captured by Sharp [70]. These were captured using a digital camera and utilising capture techniques discussed in Montgomery [67]. These images originally consisted of a sequence of images of which a subset was selected to generate a disparity map for a particular pair using StereoMatch. The images are 644*514 pixels in size, rectified but have no associated groundtruth or other generated disparity maps. In communications with Sharp [49], it was established that these stereo images contain slight variations in the intensity level depicted between the left

and right images due to slight changes in capture conditions between the image sequences. This difference may pose the system a challenge to determine if it is able to detect the differences in intensity level and highlight these within its resulting analysis.

One stereo image (As) was selected from a scan of an original stereogram image. The capturing techniques used for this pair are unknown and it is not possible to ascertain if this stereo image is rectified. The age of the original image suggests that the pair would have been captured either with a stereo camera or with no set capture technique. Hence, it would be expected that the image would contain many problems including vertical disparity. Since the images were scanned into a digital format, there will also be noise present due to this process. This image is used as part of the evaluation of the system to determine how the system behaves when presented with an image which does not conform to all the analysis requirements (non-rectified).

The final stereo image (Flower) was selected from a range available for download from the Internet. No information regarding the capture of this image is available and hence nothing can be assumed regarding the quality of its capture. Since most home users will not utilise stereo image capturing techniques for the generation of stereo images it would be anticipated that this image would be similar to those viewed by normal users i.e. there will be no data regarding the original camera calibration. The ability of the system to analyse such images is important in establishing its usefulness as a tool to aid users in assessing if the viewing quality of a stereo image is appropriate.

Table 6.1 summaries the images selected for analysis and thumbnails of the selected images are depicted in Appendix E, Figures E.1 - E.12.

In addition to the afore mentioned images, the Small Middlebury images: Barn1, Bull and Sawtooth and the Large Middlebury images: Cones and Teddy were also tested with disparity maps generated via StereoMatch. These disparity maps have been generated to establish an approximate error factor

No.	Image Type	Known Capture?	Scan or Digital?	Map available?	Size	Image Format	Scale Factor
6	Middlebury Small Images	Yes	Digital	Yes	430*380	PPM	$\frac{1}{8}$
3	Middlebury Small Images	Yes	Digital	Yes	430*380	JPG	$\frac{1}{8}$
2	Middlebury Large Images	Yes	Digital	Yes	450*375	PPM	$\frac{1}{4}$
2	Middlebury Large Images	Yes	Digital	Yes	900*750	PPM	$\frac{1}{2}$
2	Sharp Images	Yes	Digital	No	644 * 514	PNG	$\frac{1}{4}$ & $\frac{1}{2}$
1	Scanned Stereogram	Unknown	Scan	No	427*473	PNG	$\frac{1}{8}$
1	Downloaded Image	Unknown	Unknown	No	250*334 & 250*281	JPG	$\frac{1}{8}$

Table 6.1: Test Images

associated with the generation of disparity maps using StereoMatch for disparity violation and intensity difference detection. This error factor may be used to refine the disparity violation results produced by the system to account for errors in disparity map generation.

The intensity differences reported by the system when analysing stereo images constructed using JPG images is also calculated. This is based on using groundtruth disparity maps only, since the analysis of StereoMatch generated disparity maps would introduce factors associated with both the disparity maps generation, the generation of disparity maps in StereoMatch from JPG images and differences in the disparity violations detected. This analysis will give an indication of the differences caused by JPG compression in images. No analysis for disparity violation is performed as part of this testing, since the level of detected disparity is disparity map dependant and is not associated with the individual images beyond disparity map generation.

The calculated error factors associated with the use of StereoMatch generated disparity maps and JPGs as opposed to PPM images can be examined and their values considered when examining similar images. This means that it may be possible to incorporate these factors into refinements of the system to increase the accuracy of the calculations provided and effectively aid in re-evaluating and improving the system results.

6.2 Results

The results from analysing the data produced from the system analysis of the 14 test images can be found in Appendices F-H. In summary, the following were examined:

- Non-weighted disparity analysis for all test images using all relevant metrics reporting on the percentage of violation for the total, crossed and uncrossed disparity violation. The small and large Middlebury images

use associated groundtruth disparity maps the other images use StereoMatch generated disparity maps.

- Non-weighted disparity analysis for all segmented test images using all relevant metrics reporting on the average, variance, range, median and skew of the segmentation data. The small and large Middlebury images utilise associated groundtruth disparity maps and the other images utilise StereoMatch generated disparity maps.

- Weighted disparity analysis for segmented SmallCones, SmallTeddy, Cones and Sharp Farm test images using all relevant metrics reporting on the average, variance, range, median and skew of the segmentation data. The small and large Middlebury images utilise associated groundtruth disparity maps and the Sharp Farm image utilise a StereoMatch generated disparity map.

- Non-weighted disparity analysis for small and large Middlebury images using StereoMatch generated disparity maps and examining all relevant metrics. The results provide the total, crossed and uncrossed violation percentages.

- Non-weighted disparity analysis for the small and large Middlebury images using all relevant metrics reporting on the average, variance, range, median and skew of the segmentation data.

- Intensity difference analysis for all test images reporting on the average and the red, green and blue channel detected differences. The small and large Middlebury images utilise associated groundtruth disparity maps and the other images utilise a StereoMatch generated disparity map.

- Intensity difference analysis for all segmented test images reporting on the average and red, green and blue channels stating for each the average, variance, range, median and skew. The small and large Middlebury images utilise associated groundtruth disparity maps and the other images utilise StereoMatch generated disparity maps.

- Intensity difference analysis for small and large segmented Middlebury images using StereoMatch generated disparity maps reporting on the average and red, green and blue channels stating for each the average, variance, range, median and skew.

- Intensity difference analysis for small and large Middlebury images using groundtruth disparity maps with the left and right images in a JPG format reporting on the average and red, green and blue channels stating for each the average, variance, range, median and skew.

- Frame cancellation analysis for both the left and right side of the image. The small and large Middlebury images utilise associated groundtruth disparity maps and the other images utilise a StereoMatch generated disparity map.

It is possible that the average results calculated for both disparity violation and intensity level difference differ for the single and segmented images. This is due to the averaging performed for the image segments. For example, if an image consists of two segments: 100*100 and 100*20 pixels with the 100*100 pixel segment contains 500 pixels violating a given limit and the 100*20 segment contains 200 pixels violating a given limit. The 100*100 segment would have a calculated total violation of $\frac{500}{10000} = 5\%$ and the 100*20 pixel segment would have a calculated total violation of $\frac{200}{2000} = 10\%$. This would give an average percentage for the segmented images of $\frac{10+5}{2} = 7.5\%$. The single image would exhibit 700 pixels violating the limit for an image which is 100*120. This would give a calculated violation level of $\frac{700}{12000} = 5.83\%$.

6.2.1 Examination of the Disparity Violations

The analysis of the disparity violations present in the test images is provided in Appendix F. This analysis summarises the statistical results calculated for both the single and segmented test images.

An Overview of the Images

This section will examine the test images summarising the results obtained for disparity violation analysis.

Examining the Disparity Level Violations in the Small Middlebury Images

The analysis of the disparity violation results obtained for the small Middlebury images can be found in Appendix F. As discussed in Chapter 5 stereo images with disparity maps with a $\frac{1}{8}$ scale factor associated with them can only exhibit disparity violations for the Siegel and Nagata [92] limit. The small Middlebury images have disparity maps with a scale factor of $\frac{1}{8}$. This means that the only violation possible is the Siegel and Nagata (S+N) limit. This determines a limit on microstereopsis only.

The groundtruth single images exhibited a range of 20.24-55.00% of total violation of the S+N limit with the average being 35.99%. In 5 out of the 6 images greater levels of crossed disparity violation was discovered than uncrossed disparity. Examination of the statistics calculated for the segmented images demonstrated that large variations exist between segments throughout the images. For 5 out of the 6 images a variance exceeding 1000 was recorded. The total average disparity violation for the segmented images was 35.99% with a range of 18.73-53.78% for the examined images.

Examination of the image segments for each of the images demonstrated large variations for the S+N limit violation throughout. Some images exhibited segments with 0% violation and 100% in the same image and in some cases in segments adjacent to each another. Violation of this limit was present in all of the small Middlebury images where a range of at least 65% violation was recorded between the segments.

Examination of the crossed and uncrossed disparity violations in the segmented images did not provide any correlations between the segments in respect to the detected crossed and detected uncrossed disparity. This means that the crossed

and uncrossed disparity is evenly spread throughout with low/high levels of crossed violation not implying low/high levels of uncrossed violation and visa versa.

Examining the Disparity Violations in the Large Middlebury Images

The examined large Middlebury images include SmallCones, Cones, SmallTeddy and Teddy images. These images have associated disparity maps of $\frac{1}{4}$ and $\frac{1}{2}$. Both these scale factors may exhibit disparity level violations of the Yeh and Silverstein (Y+S), Siegel and Nagata (S+N) and Lipton limit.

Examination of the SmallCones and SmallTeddy single images demonstrated an average 9.05% violation for the S+N limit, 12.65% for the Y+S limit and 6.29% for the Lipton limit. The SmallTeddy image showed minimal violation in all three limits with a violation of 1.90% for the Y+S limit, 3.44% for the S+N limit and 1.12% for the Lipton limit. In both images the disparity violation was primarily caused by crossed disparity.

Examination of the SmallCones and SmallTeddy segmented images revealed an average of 8.49% total violation for the S+N limit. This is 4 times less than that detected for the small Middlebury images. This difference could be due to the increase in scale factor resulting in the range of possible disparities being increased and hence the probability of disparity found in the S+N limit range reduced. However, the SmallTeddy image exhibited 2.93% disparity violation which is a lower value than would be expected due to scale factor difference alone and hence it is likely that low violation rates are present for this image.

The SmallCones image segments demonstrated large ranges of total disparity violation throughout. The majority of the image segments exceeding 60% violation are located in the bottom and left segments of the image. A high variance is calculated for the segmented data.

The SmallTeddy image exhibited limited violations for all of the limits with a small range and small average disparity violation detected for all examined

limits. This implies that the image contains small violations of disparity and does not contain regions where violations are high.

Examination of the Cones and Teddy images revealed that a smaller S+N limit violation was exhibited in these image than for the SmallCones, SmallTeddy and the small Middlebury images. This behaviour is expected since the accompanying disparity maps exhibit a larger range of disparity thus the predicted proportion of the map denoted by the S+N limit is reduced. The behaviour of the Y+S and Lipton limits for the Cones and Teddy image appear to be similar to that exhibited in the other large Middlebury images.

Examination of the Cones and Teddy single images demonstrated an average 3.55% violation for the S+N limit, 48.59% for the Y+S limit and 38.16% for the Lipton limit. Comparisons to the small Middlebury images, the amount of violation of the S+N limit exhibited average disparity violations much smaller than the percentages exhibited by the small Middlebury images, with Teddy (3.44% single and 0.43% in segmented) and Cones (7.65% single and 6.67% in segmented). Examination of the variance and skew of the image segments demonstrated that the violation of the S+N limit is skewed towards the lower end of the range.

Examination of the results when applying a weighting function for the images SmallTeddy, SmallCones, Teddy and Cones for the total average calculated disparity violation was 23.97% for the Y+S limit and 15.86% for the Lipton limit. The SmallTeddy image exhibited low levels of disparity violation for both the Y+S and Lipton limit being 2.94% and 0.379% respectively. The other three images contained higher levels of detected disparity violation.

Examining the Disparity Level Violations in the Sharp Images

The single Sharp images with StereoMatch generated disparity maps exhibited an average of 41.31% for the Y+S limit, 23.05% for the S+N limit and 31.86% for the Lipton limit.

Examination of the statistics calculated for the segmented images demonstrated that large variations were present between segments throughout the images for the Y+S and Lipton limit. The S+N limit for the Hand image shows very minimal variance for both crossed and uncrossed and for the Farm image minimal variance is exhibited for crossed disparity however large variations are present for uncrossed disparity.

Examining the Disparity Violations in the Miscellaneous Images

The single Miscellaneous images with StereoMatch generated disparity map exhibited violations of the S+N limit only. The As image (scanned stereogram) contained an average of 10.75% violation and the Flower image (downloaded) contained an average of 16.87%. Examination of the statistics calculated for the segmented images demonstrated that minimal variation was present between the segments of the image for both of the images.

Disparity Violations Detected for StereoMatch Generated Disparity Maps

Analysis of the small and large Middlebury images for the disparity violation detected by the system when utilising StereoMatch generated disparity maps is summarised in Appendices F - H.

Examination of the small Middlebury images resulted in a detected range of 8.49 - 50.52% of total violation of the S+N limit with the average being 26.07%. In 5 out of the 6 images greater levels of uncrossed disparity violation was detected than crossed disparity. Examination of the statistics calculated for the segmented images demonstrated that large variations exist between segments throughout the images. This is demonstrated by large variances for all images. For 3 out of the 6 images a variance exceeding 1000 was recorded. The total average disparity violation for the segmented images was 27.23% with a range of 9.57 - 53.64%.

Examination of the large Middlebury images demonstrated that the SmallCones and SmallTeddy images exhibited an average 24.56% violation of the

S+N limit, 12.99% of the Y+S limit and 5.67% of the Lipton limit. The Cones and Teddy images exhibited an average disparity violation of 59.34% for the Y+S limit, 14.79% for the S+N limit and 51.92% for the Lipton limit. Examination of the statistics calculated for the segmented images demonstrated that large variations exist between segments throughout the images.

6.2.2 Examination of the Intensity Differences

Analysis of the differences in the intensity level detected by the system for groundtruth disparity maps are summarised in Appendix G.

An Overview of the Images

Intensity Difference - Small Middlebury

The average intensity difference detected by the system for the single small Middlebury images was 2.23%. Examining the red, green and blue channels demonstrated that for all 6 images the green channel exhibited the least amount of intensity difference and in 4 out of 6 of the images the blue channel exhibited the greatest amount of difference. The average for each channel was calculated to be 2.26% for the red channel, 1.93% for the green channel and 2.43% for the blue channel. Examination of the segmented images determined that there was minimal variation between the segments with the largest segment exhibiting 8.70% average difference. The largest values were found in the Sawtooth image in all three channels.

Intensity Difference - Large Middlebury

The average intensity difference detected by the system for the large Middlebury images was 3.36%. Examining the red, green and blue channels demonstrated that for 3 out of the 4 images the red channel exhibited the least amount of detected intensity difference. The average for each channel was calculated to be 3.20% for the red channel, 3.28% for the green channel and 3.58% for the blue channel. Examination of the segmented images determined

that there was minimal variation between the segments with the largest segment exhibiting 13.21% difference. The largest values were found in the Teddy image in all three channels.

Intensity Difference - Sharp Images

The average intensity difference detected by the system for the Sharp images was 1.74%. Examining the red, green and blue channels demonstrated that in both images the intensity level difference exhibited by each colour channel was approximately the same. The average for each channel was calculated to be 1.72% for the red channel, 1.72% for the green channel and 1.78% for blue. Examination of the segmented images determined that there was some variation between the segments with the largest detected difference being present at the edge of both the images.

Intensity Difference - Miscellaneous Images

The intensity differences detected by the system for the single Miscellaneous images was 3.17%. For the As image and 2.27% for the Flower image. Examining the red, green and blue channels demonstrated that in both images the intensity level difference exhibited by each colour channel was approximately the same. The maximum segment intensity difference was measured to be an average of 7.03% for the As image and 11.74% for the Flower image.

Intensity Differences - StereoMatch Generated Disparity Maps

Examination of the large Middlebury images demonstrated a total of 2.79% average intensity difference detected by the system. Examining the red, green and blue channels demonstrated that for all the images the green channel exhibited the least amount of intensity difference. The average for each channel was calculated to be 2.82% for the red channel, 2.52% for the green channel and 3.04% for the blue channel.

Intensity Differences - JPEG Images

Examination of the small and large Middlebury images utilising groundtruth disparity maps demonstrated a total intensity difference detected by the system of 2.45% for the small Middlebury images and 3.42% for the large Middlebury images. Examining the red, green and blue channels for the small Middlebury images demonstrated that for all 6 images the green channel exhibited the least amount of intensity difference with the red and blue channels being comparable.

6.2.3 Examination of the Detected Frame Cancellation

Examination of the small Middlebury images determined them to exhibit frame cancellation in all images. The average overall image violation was 36.36%. In 4 out of the 6 images examined, more pixels exhibiting frame cancellation were determined to be located on the left side of the image.

Examination of the large Middlebury images determined them to exhibit frame cancellation in all images for both image sides. The average overall image violation was calculated to be 61.05%.

The Sharp images exhibit 41.10% frame cancellation in the Farm image and 64.82% frame cancellation in the Hand image. The miscellaneous images exhibit 28.46% for the As image and 6.42% in the Flower image.

6.3 Comparing Images

This section will discuss some of the relationships discovered from the examination of the 14 test images via analysis by the system.

6.3.1 How does JPEG compression effect the detected intensity level differences?

Comparing the results of the JPEG images and to original PPM images when considering groundtruth disparity maps, resulted in a slight difference in the detected intensity difference. In all cases, except for the large Middlebury Teddy image, a slight increase in the detected intensity difference was calculated. Tables 6.2 and 6.3 summarise the differences detected for both the small and large Middlebury images. A negative value signifies a reduction from the original PPM images to the JPEG images and a positive value an increase.

The calculated average intensity difference for the images resulted in a 0.22% increase for the small Middlebury images and a difference of 0.06% increased for the large Middlebury images when comparing JPEG to PPM images.

Image	Difference (%)
Barn1	0.40
Barn2	0.00
Bull	0.12
Poster	0.49
Sawtooth	0.10
Venus	0.20

Table 6.2: Comparing Intensity Difference between JPEG and PPM images - Small Middlebury

6.3.2 How does disparity map generation with StereoMatch effect the detected disparity violations and intensity differences?

The use of the StereoMatch generated disparity maps in comparison to groundtruth disparity maps was found to considerably alter the resulting disparity violation

Image	Difference (%)
SmallCones	0.22
SmallTeddy	0.09
Cones	0.00
Teddy	-0.07

Table 6.3: Comparing Intensity Difference between JPEG and PPM images - Large Middlebury

reported by the system. An example of this is the SmallTeddy image. In the groundtruth analysis this image, it was found to exhibit minimal violation in all three of the possible limit violations (Y+S, S+N and Lipton). The detected violation being 1.90% for the Y+S limit, 3.44% for the S+N limit and 1.12% for the Lipton limit. When the image is analysed using a StereoMatch generated disparity map the violation levels reported are 15.22% for the Y+S limit, 31.32% for the S+N limit and 6.55% for the Lipton limit. This is a significant reported increase for all limits.

The overall change in the disparity violations exhibited for the small and large Middlebury images are summarised in Table 6.4 for the small Middlebury images and Table 6.5 for the large Middlebury images. A negative value signifies a reduction from the groundtruth to the StereoMatch generated disparity map and a positive value an increase. (Note that the small Middlebury images only have a comparison for the S+N limit due to no other limit violations being possible)

The differences detected between the images demonstrated that for the small Middlebury images for the S+N limit there was a reduction in the detected disparity violation in 5 out of the 6 images. On average the reduction was 12.46% for the 5 images and an average of 9.93% reduction for all 6 images. The large Middlebury images demonstrated increases in the detected disparity

Image	S+N Difference (%)
Barn1	-29.79
Barn2	-12.75
Bull	-1.45
Poster	-4.38
Sawtooth	2.71
Venus	-13.91

Table 6.4: Comparing Disparity Violations for Groundtruth and StereoMatch Generated Disparity Maps - Small Middlebury

Image	Y+S Difference (%)	S+N Difference (%)	Lipton Difference (%)
SmallCones	-12.64	3.13	-6.68
SmallTeddy	13.32	27.89	5.43
Cones	19.606	2.07	21.19
Teddy	21.025	9.33	21.02

Table 6.5: Comparing Disparity Violations for Groundtruth and StereoMatch Generated Disparity Maps - Large Middlebury

for all images and all limits except SmallCones. This image showed a reduction in the detected disparity for the Y+S and Lipton limit but an increase for the S+N limit. On average the difference was an increase of 10.33% for the Y+S limit, 10.61% for the S+N limit and 10.24% for the Lipton image.

The results demonstrate that the use of StereoMatch disparity maps alters the detected disparity violation levels considerably and hence could be an important part of the system modelling. However, the disparity results are inconclusive, due to the small and large Middlebury images providing contradictory results. It would be necessary to examine the disparity violation charges in more detail to provide a valid conclusion of the expected differences. This may prove invaluable in providing a system that is useful in its assessment of disparity level violations.

The use of the StereoMatch disparity maps in comparison to groundtruth maps appears to have minimal effect regarding the intensity difference detected by

the system for the images examined. Although, to draw a more valid conclusion it would be necessary to examine further images. The total change in the detected intensity level differences for the large Middlebury images are summarised in Table 6.6. A negative value signifies a reduction from the groundtruth to the StereoMatch generated disparity map and a positive value an increase.

Image	Total Difference (%)
SmallCones	0.49
SmallTeddy	-0.41
Cones	-0.78
Teddy	-0.59

Table 6.6: Comparing Intensity Differences for Groundtruth and StereoMatch Generated Disparity Map - Large Middlebury

6.3.3 Do correlations exist between measured crossed and uncrossed disparity violation areas?

Comparing the crossed and uncrossed percentage differences for the groundtruth and StereoMatch generated disparity maps suggested there were no correlations between the percentages of crossed and uncrossed disparity detected for an image. Segments exhibiting high or low percentage differences in the crossed analysis did not imply that the segments in the uncrossed analysis would result in the same order of result and visa versa.

Analysis of the single images did not show any correlations between measured crossed and uncrossed disparity violation measurements for the examined images.

6.3.4 How does the application of a weighting function affect the resulting disparity violation analysis?

Examination of the difference made by the application of a weighting function for the tested large Middlebury and Sharp images revealed that in general the weighted segmented images exhibit the same patterns of violations as the non-weighted images. That is, segments with large violations in the non-weighted analysis tended to also have large violations in the weighted results. The results demonstrate that in general, when the violation level of the segment is low, the difference caused by the application of the weighting is minimal. However, segments with large reported levels of disparity violation often resulted in a decrease in the percentage reported.

Using weightings in determining disparity violations can aid establishing the extent of the disparity violations present. When used in conjunction with the results of non-weighted analysis it is possible to highlight areas of the image where the violations are minimal and areas where violations are large.

6.3.5 Do relationships exist between the disparity violation limits?

Examination of the large Middlebury and Sharp images for the Y+S and Lipton limit demonstrate that there is a positive correlation between the results obtained for each. A large violation in the exhibited range of percentage violation in the Y+S limit was likely to result in a large violation in the exhibited range in the Lipton limit. Due to the Lipton limit being larger than the Y+S limit using them both in conjunction seemed to aid the assessment of the level of violation present. For example, in the SmallCones image one segment exhibits a violation percentage of 89% for the Y+S limit, however, the same segment exhibits a 44% violation percentage for the Lipton limit. From this it is possible to deduce that the difference in the violation percentage is caused by a proportion of pixels that violate the Y+S limit but not the Lipton limit.

Effectively, use of both limits can aid assessing the size of the violations present in the image.

Comparing the S+N limit, which denotes a limit on microstereopsis, with the excess disparity limits (Y+S and Lipton) demonstrated that there was a slight correlation between the two for the examined images. The results demonstrate that excess disparity violation of the Y+S and/or Lipton limit usually resulted in a reduction of the disparity violation detected for the S+N limit. It is proposed that this is due to two reasons:

1. The more pixels engaged in excess disparity violations the less will be available to exhibit microstereopsis violations.

2. Images with violations of the Y+S and Lipton limits must have a disparity map associated with them with a scale factor enabling the Y+S and Lipton limit to become violated. This means probabilistically there are less pixels which will be located in the S+N limit range assuming a disparity map with an equal spread of disparity levels.

6.3.6 Do images of increased size have comparable disparity violations and/or intensity level differences?

The Large Middlebury images (SmallCones and Cones; and the SmallTeddy and Teddy images) are images of the same scene but different sizes and with disparity maps with difference scale factors. The small images are 450 * 375 pixels and the large images 900 * 750 pixels. They have disparity maps with scale factors of $\frac{1}{4}$ and $\frac{1}{2}$ respectively. The average disparity violations detected for these images are summarised in Table 6.7. For the Y+S and Lipton limits, in both sets of images the detected disparity violations increases in the large image. For the Teddy images this involves an increase from 1.05 to 51.99% for the Y+S and 0.13 to 35.21% for the Lipton limit. For the S+N limit there is a decrease in both images.

	Y+S Difference (%)	S+N Difference (%)	Lipton Difference (%)
SmallCones	31.14	14.05	17.54
Cones	45.19	6.67	41.11
SmallTeddy	1.05	2.93	0.13
Teddy	51.99	0.43	35.21

Table 6.7: Large Middlebury Images - Disparity Violations

6.3.7 Are there relationships between frame cancellation and disparity violation?

Table 6.8 states the percentage of frame cancellation determined in each of the images and the crossed disparity violations detected for the Y+S and Lipton limit. From the results there does not appear to be a correlation between the amount of frame cancellation detected and the amount of crossed disparity depicted in the images. For example the SmallTeddy image contains 72.05% of frame cancellation but only contains 1.90% Y+S and 1.12% Lipton disparity violation.

	Total Frame Cancellation	Crossed Disparity Violation (Y+S)	Crossed Disparity Violation (Lipton)
SmallCones	48.89	23.39	11.45
Cones	49.02	43.48	34.91
SmallTeddy	72.05	1.90	1.12
Teddy	74.26	41.68	26.70
Farm	41.10	39.04	26.04
Hand	64.82	43.58	37.68

Table 6.8: Frame Cancellation and Disparity Violations

6.3.8 Natural Occurring Intensity Difference

Due to the differences in the perspective of the individual images of a stereo pair, it would be expected that there would be slight differences in intensity

levels naturally present in the stereo pair. This would be caused by both occlusion areas and difference in lighting conditions.

Examination of the large Middlebury images determined them to contain intensity differences when comparisons were made between the images using groundtruth disparity maps. The average intensity difference for the images was 3.20% for the red channel, 3.28% for the green channel and 3.58% for the blue channel. This gives a total average of 3.35%.

Examination of the small Middlebury images determined them to contain an average intensity difference of 2.26% for the red channel, 1.93% for the green channel and 2.43% for the blue channel. This gives a total average of 2.21%.

Using the results for both the small and large Middlebury images gives an average intensity differences of 2.73% for the red channel, 2.61% for the green channel and 2.86% for the blue channel. This equates to a total average of 2.78%.

Since a groundtruth disparity map is the closest representation to the actual disparity present in a stereo image the detected intensity differences will be due to both the natural occurring intensity differences in the images and the inaccuracies in determination of the groundtruth disparity maps. Effectively, the average percentage difference can be used as a baseline for the expected intensity difference of a stereo image. Although, to establish this to be a workable aspect of the model it would be necessary to examine further images to increase the accuracy of the average expected value.

6.4 Summary

The 14 test images were used to test the developed model and system included a selection of images with a range of $\frac{1}{8}$ - $\frac{1}{2}$ for their disparity map scale factors, a range of capture methods (controlled digital techniques, scanned and

unknown), a range of sizes (250*281 - 900*750 pixels) and a range of image formats (PPM, JPG, PNG). The small and large Middlebury images included in the image selection are accompanied by groundtruth disparity maps.

These images were analysed to establish their disparity violation, intensity differences and the presences of frame cancellation. These values were reported by the system as percentages for each of the detection areas. Comparisons were made using the results of the examined images to attempt to answer questions relating to the images. These questions included the effects of JPEG compression, the differences caused by the use of StereoMatch as a disparity map generator, the benefits of using a weighting function, relationships within each of the detection areas, image size differences and relationships between the detection areas.

Chapter 7

Conclusions, Critical Analysis and Further Work

This chapter will provide conclusions based on the results and analysis performed in Chapter 6. The chapter will critically analyse the devised stereo quality model and the implementation produced from it to identify weaknesses and possible problems. It will comment on how the identified issues may influence the results achieved the conclusions derived. Suggestions will be made regarding possible improvements that could provide future enhancements to the model and/or implementation. Lastly, the chapter will summarise the achievements made regarding the original aim criteria proposed at the onset of the thesis.

7.1 Conclusions

7.1.1 Assessing the Quality of the Test Images

In Chapter 6, 14 stereo images were evaluated and the results analysed for the presence of disparity level violations, intensity level differences and frame cancellation.

A summary of the results obtained from the analysis of the images and the implication on their quality level is summarised within this section. The analysis assumes that the limits used by the system are accurate. Differences caused by inaccurate limits will be discussed further in a critical analysis of the model and system in the later sections of this chapter.

The Small Middlebury Images

The small Middlebury images did not exhibit any violations of any of the 5 upper disparity limits. This means that the images do not contain excess levels of disparity and hence the disparity depicted in them should be comfortably fusible by observers with functional binocular vision. However, the images exhibit violations of the S+N limit which denotes a limit on microstereopsis. This means that some areas of the images may not be observable in stereo as the disparity level is below that of human fusibility. On average the images exhibited a 36% violation of this limit implying that an average of 36% of the pixels were below the required threshold for stereo perception. Examining the segmented images demonstrated that some areas of the images exhibited 100% violation of the limit. In such cases, for $\frac{2}{3}rds$ of observers, this may mean that the image sections are observed on the display plane and not as projected disparity. This could cause a reduction in the stereo effect of the images.

The intensity level differences detected for the images did not exceed 5% for the average difference detected. The Sawtooth and Poster image contained some image segments with intensity differences in excess of 5%, however, this

level of difference was detected in few segments. This may mean that the intensity difference is detectable in certain locations of the image.

The small Middlebury images were captured using controlled techniques and have a groundtruth disparity map associated with them. This means that the detected intensity level differences can be used to represent the naturally occurring differences in images due to the change of perspective. This is discussed in Chapter 6. The calculated results give a 2.78% benchmark value being detected due to natural light changes.

For all of the images, frame cancellation was detected as being a possible problem. There was a large variation throughout the images with regard to the percentage of frame cancellation detected.

Large Middlebury Images

The large Middlebury images exhibited violations of the Y+S and Lipton limits as well as the S+N limit. This means that the images include excess disparity levels that may be non-fusible by observers with functional binocular vision. The larger images (Cones and Teddy) exhibited larger violation levels for both the Y+S and Lipton limits (48.59% and 38.16%) than the small images (SmallCones and SmallTeddy) exhibited (12.65% and 6.29%). For the Y+S and Lipton limits all of the large Middlebury images exhibit large variation between sections of the images with each image containing segments nearing 100% violation. This means that the violation of the limits may be primarily centred in particular regions of the image and not necessarily spread throughout. The violation of the S+N limit found to be an average of 9.05% for the small images and 3.55% for the large images. These values are noticeably smaller than those found for the small Middlebury images. Examining the segmented images demonstrated that the variance was also reduced demonstrating that the variation found in the images was less than in the small Middlebury images.

The intensity level differences detected for the images did not exceed 5% for

the average difference detected. All of the large Middlebury images contained segments with intensity level differences that exceeded 5% however this level of difference was only detected in a few of the segments. This may mean that the intensity difference is detectable in certain locations of the image. However, it should not influence the depth perception of the scene.

The amount of frame cancellation detected in the images was an average of 61.05%. In all of the images frame cancellation was detected as being a problem in the image.

Sharp Images

The Sharp images exhibit violations in two of the upper disparity limits. This means that the images include excess disparity levels that may not be fusible by observers with functional binocular vision. The single images with a StereoMatch generated disparity map exhibited an average of 41.31% for the Y+S limit, 23.05% for the S+N limit and 31.86% for the Lipton limit. The Farm and Hand images compared very similarly for both the Y+S and Lipton limits with the Farm image exhibiting 39.04% and 26.04%; and the Hand image exhibiting 43.58% and 37.68% respectively. The disparity violations found for the Y+S and Lipton limits for each image were similar. Examining the segmented images demonstrated variation throughout the images especially for the Y+S and Lipton limits.

The average measured intensity differences detected was 1.74%. The measurements made within Chapter 6 suggested that a natural variation of 2.78% would be expected to occur within the images. However, using StereoMatch as a disparity map generator on average would reduce this by -0.32%. This would give an expected natural variation of 2.46%. A measured average of 1.74% difference would hence imply that the examined images exhibited little intensity level difference. This was unexpected since communications with Sharp [70] implied that small levels of intensity difference were present between the images.

Miscellaneous Images

The miscellaneous images did not exhibit violations of any of the five disparity limits determining excess disparity levels. This means that the disparity depicted in the images should be comfortably fusible by observers with functional binocular vision. However, the images exhibit violations of the S+N limit which denotes a limit on microstereopsis. This was recorded as 10.75% for the As image and 16.87% for the Flowers image. This means that some areas of the images may not be observable in stereo as the disparity level is below that of human fusibility.

As discussed in Chapter 6, the expected intensity level difference detected due to natural variation is on average 2.78%. However, using StereoMatch as a disparity map generator on average would reduce this by -0.32%. This would give an expected natural variation of 2.46%. Additionally, the stereo images are in JPEG format. Calculations in Chapter 6 determined that when comparing PPM images to JPEG format an increase on average of 0.16% was present. This would give an expected intensity difference of 2.62%.

The As image had a measured intensity difference of 3.17%. This is a slightly larger value than would be expected due to natural variation and JPEG compression, however comparing it to the StereoMatch images shows it to be within the bounds of possible natural occurring intensity differences. The Flower image contained a measured intensity difference of 2.27%. This measurement is less than that which would be expected due to natural variation and JPEG compression. Hence, for both these images it is not possible to comment regarding the presence of intensity difference within the images.

7.2 Current Model and Implementation Weaknesses - A Critical Analysis

In this section identified problems in the methodology employed in both formulating a stereo quality model, its implementation and analysis will be examined.

7.2.1 Incorrect Disparity Map Determination

The disparity map generation system employed by the system is StereoMatch. StereoMatch calculates a disparity map based on a heuristic. To function correctly StereoMatch requires configuration information regarding the stereo pair to be analysed. This information includes the range of disparity levels present in the images and an appropriate scale factor. Discussed in Chapter 4 is the necessity for these variables to be established by image inspection. This leads to the following problems:

1. **Incorrect disparity range determination**
   If a selected range is too small for the disparity depicted in the stereo image this would result in a limited level of disparity present in the stereo image being determined. This would result in a disparity map containing mismatches or large numbers of undeterminable disparity levels (which may be incorrectly assumed to be occlusion areas). An example of a disparity map calculated with StereoMatch where the range is too small is given in Figure 7.1. The left image is produced with 0-10 range and the right image with 0-80 range. The scale factor is $\frac{1}{8}$ for both.

   If a selected range is too large for the disparity depicted in the image this would result in all of the disparity in the image being detectable, however, there will be a noticeable degrading in system performance especially for large images. For each pixel in the left image of the stereo pair StereoMatch would perform a search space calculation. If the size of the search space is denoted by s and the number of pixels in the

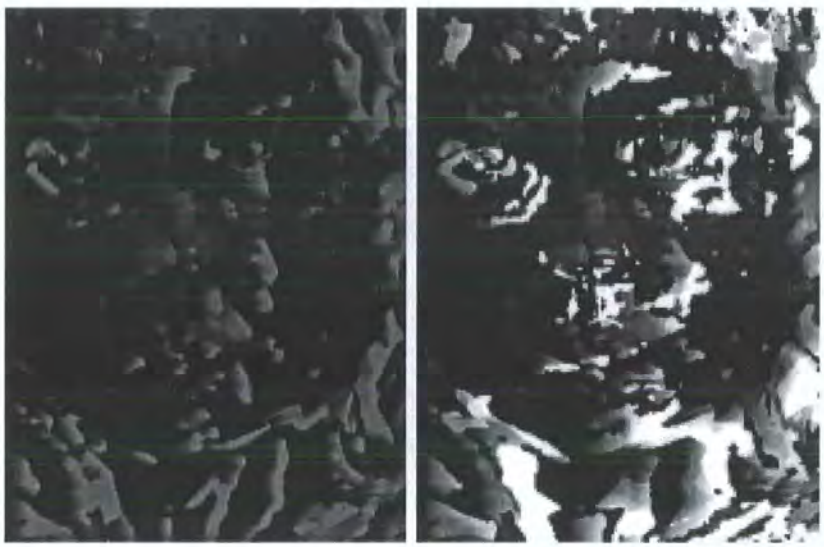


Figure 7.1: Different Ranges of Disparity (left - 0-10 and right - 0-80)

image is denoted by p the number of calculations required is given by $s * p$. Hence, increasing the search space by one pixel in size will increase the necessary calculations by p. Experiments using StereoMatch have determined that examining large search spaces especially in images exceeding 900 * 750 pixels cause a sudden degradation in performance which increases considerably as the image size becomes larger and the search spaces are extended. Analysing stereo images of size 1800 * 1500 pixels with a search space range of 0-200 pixels failed to complete its calculation within 48 hours on a dual processor and high specification memory machine.

Additionally, experiments have found that when StereoMatch is presented with large search ranges it is prone to providing false matches. This means that using the maximum permitted search space for the examined image is not a realistic solution to the configuration information problem. Hence, it is necessary to determine a suitable range manually. This process is not infallible. It is possible that the search space is estimated to be too small for the encoding present in the stereo pair and hence the resulting disparity map may contain incorrect matches or

Figure 7.2: Different Scale Factors - (left - $\frac{1}{8}$ and right - $\frac{1}{2}$)

undetermined correspondence. Additionally, it means that the implementation is not fully automated.

2. **Incorrect scale factor determination**
   Figure 7.2 demonstrates the production of two disparity maps for the Sharp Farm stereo image. The differences in the disparity map can be clearly seen between the configuration information supplied. The left image consists of a disparity map generated with a scale factor of $\frac{1}{8}$ and the right a scale factor of $\frac{1}{2}$. The range remains constant in both images.

   If a scale factor which is too large, is selected, the disparity map will be able to depict the necessary disparity but there will be a reduction in the detail the disparity map is able to depict. For example, if a scale factor of $\frac{1}{2}$ is selected for a disparity map which would be more appropriately represented by a $\frac{1}{4}$ disparity map scale factor the level of detail depicted by the map may be halved.

   If a scale factor which is too small is selected the disparity map may be unable to depict the encoded disparity between the stereo pair. This may result in the disparity map being saturated to full capacity because the actual disparity detected exceeds that permitted by the scale factor of the disparity map. This will result in an underestimation of the disparity

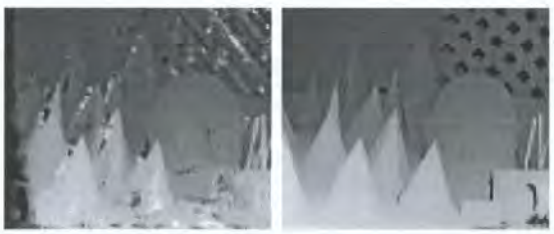


Figure 7.3: Comparing StereoMatch Generated (left) and Groundtruth Disparity Maps (right)

present in the images.

3. **Errors Due to StereoMatch**
   StereoMatch uses a heuristic to calculate a valid disparity map for a given stereo pair. StereoMatch is unable to generate good quality disparity maps in some image situations and it is likely to produce problems at edge sections (where disparity levels change) within images. It can provide false correspondences and areas of indeterminable disparity levels. This can be clearly seen when comparing the output with a corresponding groundtruth disparity map. This is demonstrated in Figure 7.3

   The errors introduced by the use of StereoMatch may propagate throughout the system since the disparity map is used extensively throughout. At the time of writing there is no known system able to generate correct disparity maps given any stereo image. This means that it is not possible to currently solve this problem.

The developed system allows users to supply their own disparity map. This means that if a groundtruth disparity map is present this can be utilised by the system and a more accurate analysis can be achieved. Additionally, if the user wishes to use other disparity map generation systems which are able to generate more accurate disparity maps, these can be used as an input to the system. This makes the system more versatile and able to adapt to changes in knowledge and disparity map generation heuristic design or possible algorithms of the future.

7.2.2 Disparity Level Limits

The human factors research employed in formulating the disparity limits in the quality model of this thesis was principally derived from human factors research performed via the use of stereoscopic display devices and stereovision experiments. Many of the experiments were performed using poor quality stereoscopic displays often suffering from large degrees of crosstalk and display images containing quality issues. These factors may have influenced the final results regarding the human tolerance levels measured. These were major problems found in display systems in the past. This means that it is questionable if the original human factor limits are still valid when applied to modern stereo display systems such as autostereoscopic displays. The introduction of modern stereoscopic display systems, specifically the development of autostereoscopic displays, means that stereo images can be presented to observers without large levels of crosstalk and image degradation.

Experimentation at the time of writing to establish accurate limits when using modern stereo display hardware have not been established. This means that it is not possible to ascertain if the original limits calculated remain accurate within the application of a model to assess disparity limits on autostereoscopic display systems.

7.2.3 Disparity Limit Violation Implementation

The implementation of a disparity limit violation calculation mechanism involves the determination of a midpoint. The method adopted for the midpoint calculation involves examining the range of disparity levels present in the disparity map. A midpoint was selected by determining the midrange level. For example, if a range of 100-242 units was found in the disparity map a midpoint of 171 would be selected.

Using this method means that the range of encoded disparity is small and the variance is minimal this method of midpoint determination could cause the original horizontal disparity encoded in a stereo pair to be lost. For example, it is possible that a disparity map reports the detected disparity throughout an image to be 255 units. Using this method a value of 255 units would be calculated for the midpoint. This would result in no disparity being depicted. Effectively, this would remove any stereo cues from the image. However, a stereo image would be of questionable benefit if it depicted horizontal disparity at one level throughout. A scene that displays disparity at one level only, would not provide any binocular disparity because the cue is based on the difference in disparity from a point of fixation to another point in the scene. Since all the points in the scene would be at the same level this would result in no binocular disparity being perceived. Hence, when the range of disparity levels present in an image are small, such images are best viewed monocularly. Additionally, monocular viewing will not include any of the other problems that are associated with stereo viewing.

Stereo images with a disparity map depicting a small variance and a large overall range will also cause a problem in the method employed to determine a suitable midpoint. For example, a disparity map with a level of 255 units throughout, except one pixel with a value of 1 unit, would result in a calculation of a midpoint of 127 units. In such an example, this would result in the majority of the stereo image being depicted as crossed disparity and one pixel in the image being depicted as uncrossed disparity. This may mean that the disparity is inappropriately displayed and the resulting viewing experience uncomfortable.

Methods for reducing incidents of inappropriate midpoint determination could be employed. The following are possible solutions that could be used to adapt the current implementation if future revisions to the implementation and model:

1. Calculate a midpoint based on the mean of the levels depicted in the disparity map. This method would give a good overall solution, however,

it is possible that some disparity is shifted towards the mean, hence the perceived disparity may be reduced for one perspective of disparity and increased for the other.

2. Calculate a midpoint where the variance is examined and the midpoint value adjusted accordingly. In the above example where the calculated variance is small, the midpoint could be adjusted towards the mean value - this method would prevent a situation where minimal pixels were present at one extreme of disparity level. However, it would be necessary to derive a suitable function for translating the variance into an appropriate change in the calculated midpoint.

7.2.4 Problems Determining Intensity Difference

The presences of occlusion in stereo images can be established from an examination of the disparity map to determine image areas where the disparity map generation system was unable to determine correspondence. However, it is possible that the disparity map generation system is unable to determine correspondence because of incorrect configuration information or heuristic limitations. This means that corresponding pixels may not always contain a match in the disparity map. Such pixels will not be included in the intensity difference analysis because there is no correspondence depicted in the disparity map. It is possible in such cases that pixels with large intensity differences are not detected and that the overall intensity difference results are affected.

Additionally, a disparity map generation system that produces correspondences with false matches may cause the detection of intensity differences where they are not present. In such situations, the intensity difference calculated will be caused by non-corresponding pixels and not from the actual intensity difference. This could mean that the intensity difference calculated by the system is incorrect and would incorporate false matches which cause greater intensity difference to be reported than that present in the actual images.

An additional problem in the determination of intensity difference occurs due to the heuristic employed in generating the disparity map. Some disparity map or correspondence generation systems produce their analysis by incorporating intensity comparisons. In some cases, such systems will use the intensity level and colour of a pixel to aid determining a matching pixel. Some systems weight the intensity level and colour of the pixel in the overall correspondence determination. This means that since the disparity map generation system incorporates intensity comparison calculations, the generated disparity map may be biased towards formulating minimal intensity differences. This may mean that the intensity level comparison produced as a product of the system does not reflect the true intensity level differences of the stereo image. This is because the differences have been removed by errors in the correspondence calculation in the disparity map generation process.

This thesis defined three limits for intensity difference present in stereo images. The Pulfrich effect was stated as occurring in the presence of intensity level differences exceeding 10%. However, the detection of intensity differences by the human visual system is not fully known. Estimates place this at approximately 5%, however, no experimental evidence has been produced to determine this quantitative difference and the size of the difference area which would cause its detection. To determine this it would be necessary to perform controlled experimentation with human observers to establish the necessary level needed by the human visual system to detect a difference between the images. Additionally, a limit for binocular rivalry of 70% was chosen as an arbitrary measurement within the thesis, however, the difference in intensity level and the size of the region required to cause this are unknown. It would be necessary to perform controlled human experimentation in order to achieve a more precise limit.

7.2.5 Problems with Determining the Presences of Frame Cancellation

Currently, computer science and psychology research is very sparse in the examination of the frame cancellation phenomenon. Little if any controlled experimentation has been performed to determine the size of the frame cancellation zone and additionally the size of its influence on binocular disparity beyond the frame cancellation area. The definition of frame cancellation produced as part of this thesis defines the frame cancellation zone to be 20 pixels to the right and left side of a stereo image. This definition was used to provide a system able to detect frame cancellation. However, it is unclear if the use of region of 20 pixels is appropriate. This is especially a concern since the horizontal size of a pixel is determined by the display and is not a set unit. From this definition it would be implied that as the pixel size increases the frame cancellation zone would also increase proportionally to it. It is unclear if this relationship occurs or if the size of the frame cancellation zone is a set screen displacement or additionally if it is influenced by other factors. Other possible factors that may influence the size of the frame cancellation zone and the perception of binocular disparity around the area are summarised in the following:

- **Display boundary size** - The size of the boundary of a display or image may effect the overall influence of frame cancellation in a stereo image. The extent of this effect or even its occurrence is currently unknown. Studies and observations, by experts in the field, have determined that it is necessary for an autostereoscopic display to contain a boundary. Suggestions are that the presence of a boundary aids observers in ascertaining the position of the display panel in space and aids them gaining a better perspective of the 3 dimensional space projected by the display. However, the size of the optimal boundary is unknown.

- **Display boundary colour and texture** - The colour of the display boundary and texture may also influence the binocular disparity levels perceived by an observer. It is unknown if this influence occurs or if

the colour and/or texture and colour of the display boundary should be adapted depending on the image displayed. However, suggestions by experts in the field suggest that the colour and texture of the display boundary may effect the viewing experiences of the display. A quantifiable analysis of this has not currently been performed.

- **Image size** - A large stereo image may be less vulnerable to frame cancellation than a smaller image. A larger observation area will give an observer a larger area to obtain a sense of binocular disparity and general depth within a scene. This means that the differences detected at the boundary may be less influential in the overall depth perception; hence the effect caused by frame cancellation may be reduced. This scenario can be related to real world observations when observing a scene through a window. The window provides the boundary to the outside surrounding area. If the window is large enough the observer will see depth as normal without too many problems caused by the window boundary. However, if the window is small in relation to the size of the scene the user may experience confusion. However, if the same window is examined where the observer is physically closer to the window this results in the scene outside of the window filling more of the observer's viewable space and the scene outside the window will become more fusible. This suggests that observing larger images makes the effects caused by the boundary less influential.

- **Display size** - The above discussed scenario can also be applied to display size. A larger display screen may reduce the influence of frame cancellation and generally make stereo images more comfortable to view. This observation is well known, meaning that manufacturers of autostereoscopic displays are progressively increasing the size of their display systems. However, as the size of the display panel is increased so is the amount of binocular disparity depicted by the display. This increase in disparity level may cause problems associated with excessive disparity in stereo images. This means that a compromise needs to be achieved. However, a valid trade-off is currently unknown.

- **Disparity levels** - The amount of disparity present at areas approaching and inside the frame cancellation zone may influence the overall effect caused by frame cancellation. For example, if the disparity is large this means that the displacement from the display panel to the point of projection will be large too. Hence, the difference at the display panel will reflect this. Small disparities mean that the displacement is also small resulting in a reduced difference. This may influence the effect caused by frame cancellation in that larger differences produce greater frame cancellation effects. However, no studies have been performed to ascertain if this is the case.

The current frame cancellation model additionally poses limits due to utilising a range of screen parallax for the frame cancellation zone. The model imposed the restriction of 0-20 pixels. This means that the full extent of an object approaching the frame cancellation zones is not always established. For example, an object that is 50 pixels horizontally and approaching the edge of the screen may be detected as being present in the frame cancellation zone, however, the size of the object causing the frame cancellation will remain unknown. The maximum area detected by the system will remain in the range 1-20 horizontal pixels. Since the effect of frame cancellation will cause a reduction in the binocular disparity perceived, it results in the effect of binocular disparity being reduced for the surrounding area. This would specifically include the object in contact with the frame cancellation zone. For example, an object containing crossed disparity present in the frame cancellation zone and extending out of the zone into the normal viewing area. The object will be perceived at a particular crossed disparity level which will then reduce at the edge to the display plane. This may reduce the overall perceived depth for the observed object. This scenario is depicted in Figure 7.4.

Additionally, the proposed method for detecting frame cancellation does not give an indication of the severity of the detected crossed disparity. It merely gives an indication of the amount of crossed disparity present in the defined frame cancellation zones. This is specifically important since it can be assumed that the larger the level of crossed disparity present the greater the effect an

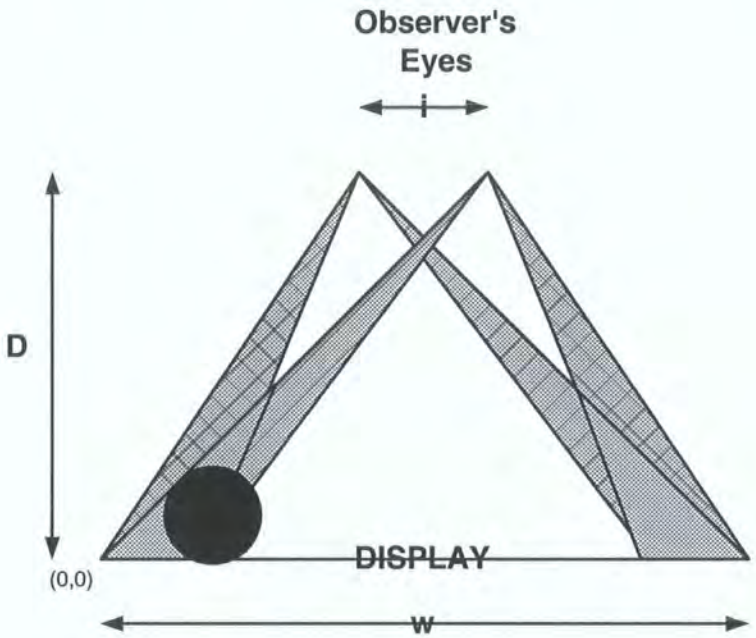


Figure 7.4: Frame Cancellation - Inside and Outside the Frame Cancellation Zone

occlusion and binocular disparity conflict will have. Detecting the extent of the crossed disparity level may aid in establishing how much influence frame cancellation may have on the image.

A method of establishing the extent of the crossed disparity present in the frame cancellation zones could be achieved using the limits utilised in the disparity detection process. This would effectively give the pixels an appropriate weighting to more accurately represent the extent of the crossed disparity present. It would be necessary to either provide six measurements for frame cancellation in a similar manner to the method employed for disparity level determination or to determine a suitable limit to be used to assess frame cancellation more accurately.

7.2.6 The Human Visual System - The Adaptation Maestro

The human visual system is very equipped to deal with many changes presented to it. This means that even in the event of scenes which are not initially acceptable it is possible that the system can adapt so that the side effects associated with their viewing are eliminated. One very good example of the adaptation power of the human visual system is experimentation performed regarding inverted scenes [29]. Adult subjects were selected for the experimentation and asked to wear spectacles that inverted their environment so that they initially observed objects at the top of the scene at the bottom of their perception and visa versa. Observers were asked to wear these spectacles for two weeks and various experiments were performed throughout their usage. It was discovered that observers initially struggled with their new inverted environment but given time they began to perceive the environment in a 'normal' fashion; in a similar way as they did prior to the experimentation. After a two-week period the subjects were asked to remove the spectacles and information was gathered regarding their perception in the natural viewing environment. Scientists discovered that when the spectacles were removed the subjects experienced confusion similar to that encountered when they initially began wearing the spectacles at the beginning of the experiment. The subjects initially observed the natural environment as inverted before their vision returned to normal. This discovery is a good demonstration of how extreme and quickly the human visual system can adapt to change in its environment.

7.2.7 Psychological Measurement Errors

A major concern regarding psychological measurements made in the literature is that the vision of subjects involved in testing should be established to determine any visual defects which may effect the results of experimentation. It is not always clear that such testing procedures have been implemented. This could result in experimentation being performed on subjects with vision defects. This is especially problematic when experiments regarding binocu-

lar vision are performed on people with unknown and/or unrectified defects in either their monocular or binocular vision. That is not to say performing experimentation on subjects with defective vision is not useful in gaining an understanding of the potential errors when defects are present in observers' vision. However, it is necessary to establish the level of a subjects vision including binocular vision assessments when performing any experimentation to ensure that results are accurate and in context.

An additional source for possible errors in the experimentation is due to the experimental testing conditions. It is apparent that in a large number experiments were performed using either unspecified apparatus or stereo display technology of 10-20 years ago in the testing procedures. This means that the experimentation has often been performed on display systems with differing quality levels of projection. This may result in any measurements being influenced by the detection of inaccuracies caused by the projection hardware and not necessarily the stereo image perception. This could lead to inaccurate results and derived image quality theories.

Other visual perception experiments have been performed in real world situations and hence the results may not be applicable in the context of autostereoscopic displays, which provide an experience different to that experienced in real world observations due to the limited imposed by the display.

The differences in the experimental conditions and the use of modern autostereoscopic displays may mean that the measured limits would be different if the experiments were repeated using autostereoscopic display hardware. This may mean that the limits used in the devised quality assessment model may be inaccurate and may need revising in the presences of more updated experimental data.

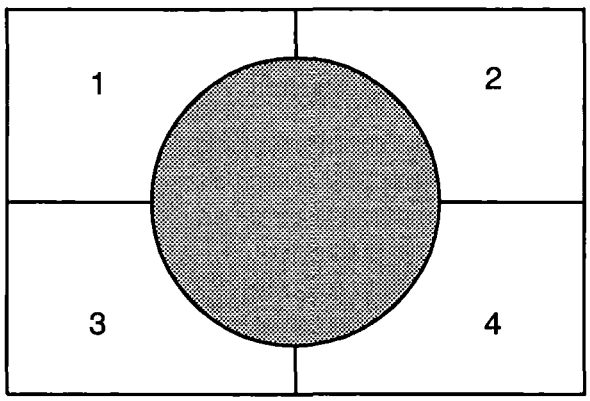


Figure 7.5: Segmentation Problems

7.2.8 Segmentation Problems

Consideration should also be given to the way in which the system performs segmentation. The system simply segments an image into 100*100 pixel sections throughout. This means that it is possible an area of an image exhibits large levels of violation of a particular limit but is not reported in the 100*100 pixel segmentation approach. An example of this is depicted in Figure 7.5. Depicted in the figure is a circular object, shaded grey, used to signify an object with a large disparity violation between segments 1, 2, 3 and 4. It would be possible that the overall percentage of violation detected within such an image and its segments would be smoothed out in the overall result meaning the violating object would not be detected as a problem within each of the segments or the entire image.

7.3 Further Work

This section details some of the areas identified for further work as a result of this thesis. This include improvements to be made to the model and implementation, human factor analysis and experimentation testing procedures.

7.3.1 Current Model and Implementation Problems

The following have been identified as possible problems in the current model and implementation:

The determination of the Scale factor and Range Variables for use by StereoMatch

The use of StereoMatch as a disparity map generation program requires the input of configuration information in relation to the stereo image. This includes both an appropriate scale factor and a suitable disparity range. An expert in the area currently performs the process of determining these variables by image inspection. This means that the implementation is currently not automated and requires human interaction. Providing a method to determine these variables automatically would solve this problem.

Disparity Limit Analysis

Experimentation to derive human factor limits for acceptable disparity levels on autostereoscopic display systems were primarily performed using poor quality stereoscopic display systems. The quality of the observer's vision and binocular vision were not stated and hence it is difficult to know if such experimentation may have been performed on observers with vision defects. This means that it is not possible to know if such limits are applicable to modern stereoscopic display systems. Analysis of disparity limits on modern displays using controlled psychological visual experimentation would allow an expansion of the accuracy of the current system with perhaps a reduction in the required number of metrics to assess the disparity violation effectively.

Intensity Difference Analysis

Experimentation to derive human factor limits for acceptable intensity difference in stereo images is sparse. Current knowledge is limited to expert's perceptions and non-quantitative definitions. Formal experimentation regard intensity level difference recognition and binocular rivalry has not been performed. The use of 5% for human recognition and 70% for binocular rivalry

were formulated from discussions with experts, however, to enable the model to be more accurate controlled human psychology experimentation should be performed and the limits.

Frame Cancellation Analysis

The model of frame cancellation was devised due to discussions with Holliman [42] regarding a 20 pixel limit. Discussed in the preceding sections of this chapter are some of the problems with choosing such a limit. To enable the model to perform frame cancellation analysis more accurately controlled human psychology experimentation should be performed to determine the limit more accurately.

Frame cancellation is defined as being the presence of crossed disparity at the left and right borders of an image, however, little attention is paid to the top and bottom of an image and how crossed disparity in these regions could effect binocular disparity perception. Additionally, frame cancellation is defined as being caused by crossed disparity only. However, logically if there is uncrossed disparity present in the frame cancellation zone this will be reduced to a perception on the screen plane. It is unknown if this would have the same effect as crossed disparity or why the phenomenon of frame cancellation is only defined as being caused by crossed disparity.

7.3.2 Further work regarding the Current Model, Software system and Testing procedure

Comparing the system analysis to human psychology experimentation

The current system determines if a stereo image is suitable for human viewing by detecting the presences of excess disparity, intensity differences and frame cancellation. Each of these identified factors are known to cause humans visual problems when observing in stereo. The developed system reports on each factor to give an overall analysis of the quality of the stereo image for human

viewing. However, it is unknown if the results are comparable to the perception of the average observer. This includes establishing if an image which is deemed as problematic by the system is perceived as such by the observer and additionally if an image that is deemed suitable for viewing by the system is also perceived as suitable by an observer.

To achieve this it would be necessary to implement controlled psychology testing. The test stereo images would be shown to observers and their perception of the image recorded. It would be necessary to perform the testing on a sample of observers representative of the general populous where visual factors and other factors which may effect an observers perception are carefully monitored.

The testing of the system and the results of psychology experimentation using a the stereo images could be compared to determine if the system results correlate with the perception of observers. This may aid in determining if the system uses an accurate stereo quality model and assessing the system's usefulness in establishing the quality of stereo images. It may also identify current restrictions on the model and areas for additional improvement. Additionally, using the comparison results it may be possible to refine the system further to provide a greater accuracy of image quality judgement or to incorporate other influential factors.

A limited number of test cases

The system was tested using 14 test stereo images. These were analysed to establish their disparity violation, intensity difference and the presence of frame cancellation. These images ranged in size, disparity map scale factor, type of capture and suitability for use in the system. They included images with groundtruth disparity maps and images where no groundtruth disparity maps or other disparity maps were available.

Six small and four large Middlebury images were used to establish the differences reported by the system for disparity violation and intensity differences when using StereoMatch as a disparity map generator in comparison

to groundtruth disparity maps. These images have disparity maps with scale factors in the range $\frac{1}{8} - \frac{1}{2}$. In order to calculate an accurate error factor when comparing groundtruth images with those generated via StereoMatch it would be necessary to examine a wide range of images including size, disparity map scale factor and capture conditions. However, it would be necessary to have a large selection of stereo images with groundtruth disparity maps available.

Expansion of the system to incorporate other processes

The thesis developed a basic model of stereo image quality and a simple implementation based on the model. It is not intended to be a commercially viable or extensively useful system in its current state. However, it may be used as a basic framework for developing a more comprehensive model and system implementation which could provide a valuable and usable tool for assessing the quality of stereo images.

The system currently assesses three areas known to affect the quality of stereo images. However, as discussed in Chapter 3 there are various identified factors that can affect the viewing quality of stereo images. Not all of these factors have been incorporated into the devised model. Factors such as determining disparity gradient violations, measuring disparity randomness and measuring intensity level variations etc. were not considered. Additionally, there are other elements of the system implementation and model that could be expanded and adapted to increase the usefulness and assessment capabilities of the system. These factors include the following:

- **Rectification**

  The system currently assumes that stereo images are rectified prior to analysis. If non-rectified images are analysed with the current system it would be anticipated that errors would be present due to the generation of a disparity map with StereoMatch and the quality analysis performed by the system. Providing the system with a method to rectify the images prior to analysis would enable the analysis of non-rectified images.

- **Disparity Randomness and Disparity Gradients**
  The system currently performs a basic assessment of the variance of disparity and intensity level violations. It performs this analysis based on the segmentation of the image and provides a basic comparison between the segments. However, it does not give consideration to how the disparity levels are distributed in the images and the disparity gradients present.

- **Disparity Map Generation**
  The disparity maps are generated by the system using StereoMatch. StereoMatch is based on a heuristic which provides disparity maps of reasonable quality when presented with stereo images adhering to certain predefined conditions. In certain image situations it is not able to provide a suitable disparity map to allow the stereo image to be effectively analysed by the system. Additionally, StereoMatch requires the provision of image parameters prior to examination. This is currently performed as a human interactive step. To improve the automation of the system and hence the usability by non-expert users it would be necessary to either utilise a different disparity map generation system or to provide the system with a means to calculate the required information.

- **Determining Mismatching Images**
  Currently the system accepts a pair of stereo images and assumes that the images have some degree of correspondence between them. This could result in the analysis of a stereo pair that does not contain a valid correspondence. In such situations it would be anticipated that the results provided by the system would indicate large levels of intensity differences. To prevent this, the introduction of a process to determine if a stereo pair is likely to be matching could be incorporated into the system. This would be beneficial to the overall image analysis.

  Currently there is no accurate solution to determining if a presented

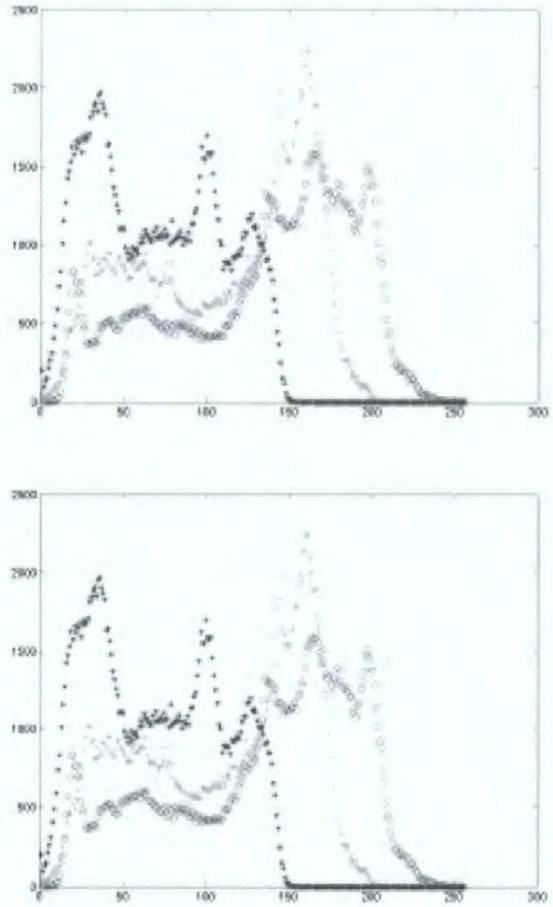


Figure 7.6: Histograms of Intensity Levels Present in Images - Left (top) and Right (bottom) - Red, green and blue channels

stereo pair is matching. However, preliminary experimentation involving the production of intensity level histograms for the left and right images of the stereo pairs revealed that the patterns exhibited by a matching pair are similar in shape, with the peaks and troughs being located in approximately the same intensity regions. This is demonstrated in Figure 7.6.

- **Improved Segmentation**
  The system current incorporates a basic form of segmentation. However,

as discussed within this chapter such a method of segmenting an image may result in areas of large violations remaining undetected. There are other possible methods that could be used to segment the images. This includes artefact segmentation. Such a method of segmenting the images may improve the ability of the system to detect large violation regions. It may also allow the identification of artefacts in the image which contain large violation of particular limits.

7.4 Success Criteria

In Chapter 1 a list of thesis aims were established. This section will revisit these aims and review how successful the thesis has been in fulfilling the success criteria.

To provide an overview of human vision including the history and current research into stereo human factor analysis.

In Chapter 2 the thesis introduced human vision including how the world is perceived both in terms of monocular and binocular perception. It then gave a detailed analysis of how humans perceive depth in scenes. Chapter 3 discussed human factors experiments, observations and discussions with experts to produce an overview of factors effecting stereo viewing.

To provide an introduction to stereo display technologies and image capture.

Chapter 2 provided an overview of stereo display technology including an analysis of the problems encountered due to their construction. The production of stereo images via the use of stereo camera hardware and techniques was overviewed and analysed.

To provide a geometric model of stereo perception on autostereoscopic display systems.

A geometrical model of stereo viewing on autostereoscopic display systems was established from the literature and presented in Chapter 2. This was used to calculate a set of equations that clearly model stereo viewing on autostereoscopic displays. The mathematical derivations of this were given in the appendices. Terms used in the literature were assessed and evaluated included highlighting definition inaccuracies.

To develop a basic model for assessing the quality of stereo images.
Chapter 4 examines human vision and human factors analysis and formulates a set of possible factors effecting quality in stereo images. The model is devised by evaluating the examined factors and formulating a subset of factors that would be appropriate for a devised quality model. The devised model is a preliminary framework for stereo image quality assessment which could be used to build a more comprehensive stereo image quality assessment evaluation criteria.

To develop an implementation of the stereo quality model able to analyse a subset of stereo images.
The system was developed from the devised model of stereo image quality and implemented in MATLAB. It uses StereoMatch to generate its disparity maps. The process is not fully automated and requires human interaction for the calculation of configure information for the use of StereoMatch. The systems performs the following implemented analysis:

- Calculation of the percentage of disparity violation for both single and segmented images. This analysis includes measurements incorporating non-weighting and weighting function methods. The result is a percentage indication of the disparity level violation for an entire stereo image and its 100*100 pixel segments.

- The matching pixel to pixel intensity differences in both the entire image and image segments. The result is a percentage indication of the intensity difference for an entire stereo image and its 100*100 pixel segments.

- The percentage of frame cancellation detected within a stereo image present at both the left and right side of a stereo image.

To evaluate the stereo quality assessment system using a set of test images.

The details of testing the devised stereo quality system using 14 stereo images of varying quality and type were provided in Chapter 6. Each test image was analysed for the presence of disparity violations, intensity differences and frame cancellation and a report of the results produced. These tests provided benchmarks to aid evaluating the model and providing possible refinements to the model and its implementation. Additionally, an evaluation of the limits and framework used to determine the disparity violation, intensity difference and frame cancellation criteria is provided and suggestions for further work considered in order to improve the current model.

The evaluation of the model and implementation including possible refinements, pattern recognition and a critical analysis.

This chapter provides a detailed evaluation of the thesis, critically analysing both the devised model and its implementation. This analysis is related to the discovered results to detail possible errors that may have influenced the overall test image analysis. Suggestions regarding improvements for both the model and implementation are also provided as well as a discussion of further work.

Appendix A

Derivation of the Disparity Equations

Equation 2.1 and Figure A.1 state how to calculate the value of s, the screen parallax. Xleft is the horizontal component of the coordinate from where the left image resides on the screen plane and Xright is the horizontal component of the coordinate from where the right image resides on the screen plane when given a display plane with the left side of the display at coordinate (0,0). This results in $s < 0$ for crossed disparity and $s > 0$ for uncrossed disparity. When this is translated into the number of pixels on a given display, $n < 0$ for crossed disparity and $n > 0$ for uncrossed disparity.

V is the distance from the observer to the point of vergence. When an observer focuses on a point projected into 3D space their eyes converge onto the point, this is known as the point of vergence. Using Figure 2.6 and the magnitude of s, equation 2.3 can be derived:

Since for crossed disparities the value of s is negative and uncrossed disparities the value of s is positive (as shown in equation 2.1) these equations can be combined to give:

Crossed Disparity	Uncrossed Disparity
$tan(a) = \frac{i}{2V} = \frac{\|s\|}{2(D-V)}$	$tan(b) = \frac{i}{2V} = \frac{\|s\|}{2(V-D)}$
$\Rightarrow \frac{i}{V} = \frac{\|s\|}{(D-V)}$	$\Rightarrow \|s\| = \frac{i(V-D)}{V}$
$\Rightarrow \|s\| = \frac{i(D-V)}{V}$	$\Rightarrow \|s\| = \frac{(iV-iD)}{V}$
$\Rightarrow V = \frac{iD}{\|s\|+i}$	$\Rightarrow V = \frac{iD}{(i-\|s\|)}$

Table A.1: Deriving V

$$V = \frac{iD}{i - s}$$

For uncrossed disparity $s < i$ otherwise the eyes would diverge and be unable to converge onto a single point in space. For crossed disparities s can either be less than or greater than the value of i and $s < 0$. This results in $V > 0$ for uncrossed and crossed disparities.

Screen depth (Sd) is the displacement of the projected point from the display screen in relation to the observer and screen. Using Figure 2.€, Equation 2.4 is derived:

$Sd = D - V$

$\Rightarrow Sd = D - \frac{iD}{i-s}$

$\Rightarrow Sd = D + \frac{iD}{s-i}$

Hence giving the equation for Sd:

$$Sd = D + \frac{iD}{s - i}$$

For uncrossed disparities $s < i$ and $s > 0$ resulting in $Sd < 0$. For crossed disparities since $s < 0$ the result is that $Sd > 0$.

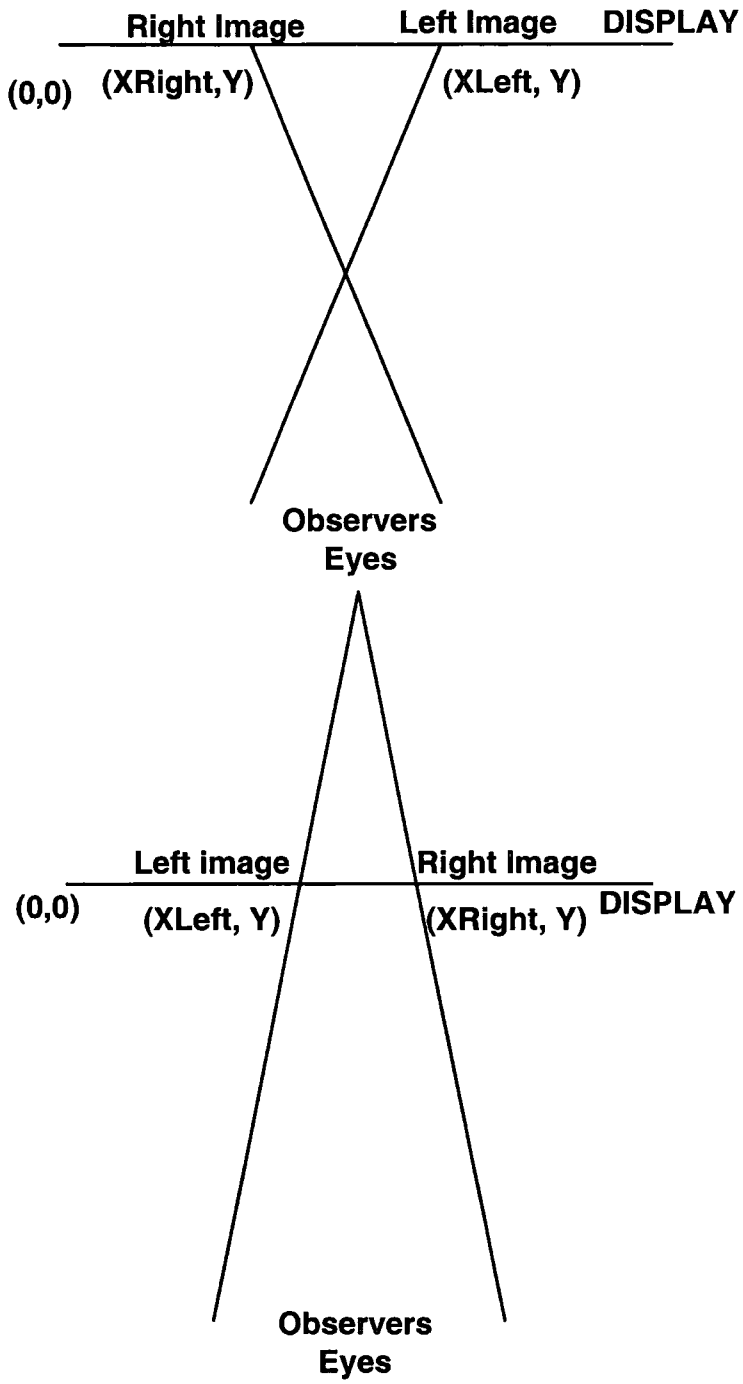


Figure A.1: Calculating s - Crossed (top) Uncrossed (bottom)

Appendix B

Vergence Difference - Resolving the Definition of the HVA

Wartell et al. [110][111], Hodges [39] and Lingard [59] all describe the Horizontal Vergence Angle (HVA) to be modelled as shown within Figure B.1. HVA can be derived using as follows:

$$\text{HVA} = 2arctan(\frac{s}{2D}) \tag{B.1}$$

It can be shown from this equation that for crossed disparity HVA < 0 and for uncrossed disparity HVA > 0.

Hodge [39] stated that the HVA is equal to vergence difference (the difference between the focus and vergence angles). However, using Figure 2.7 and the following derivations this is shown to be incorrect.

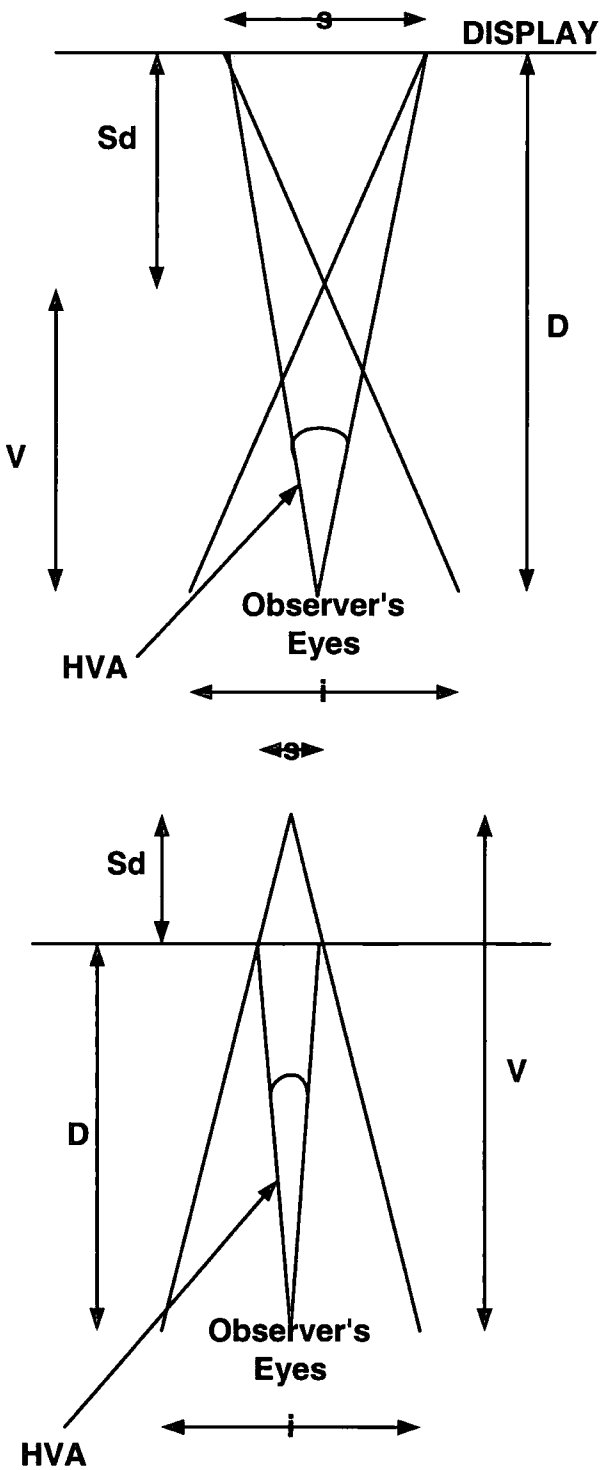


Figure B.1: HVA - Highlighted between the red lines

For crossed disparity:

$VD = b - a$

From Figure 2.7 using basic trigonometry the values of a and b can be derived:

$b = 2arctan(\frac{i}{2D})$

$a = 2arctan(\frac{i}{2V})$

Substituting in the value of V gives:

$a = 2arctan(\frac{i(i-s)}{2iD}) = 2arctan(\frac{(i-s)}{2D})$

Inserting these into the vergence difference equation gives:

$VD = 2arctan(\frac{i}{2D}) - 2arctan(\frac{(i-s)}{2D})$

For crossed disparity $s < 0$. Resulting in $2arctan\frac{i-s}{2D} > 2arctan\frac{i}{2D}$.

Therefore $VD < 0$

For uncrossed disparity:

$VD = b - c$

From figure 2.7 using basic trigonometry the values of b and c can be derived:

$b = 2arctan(\frac{i}{2D})$

$c = 2arctan(\frac{i}{2V})$

Substituting in the value of V gives:

$c = 2arctan(\frac{i(i-s)}{2iD}) = 2arctan(\frac{(i-s)}{2D})$

Inserting these into the vergence difference equation gives:

$VD = 2arctan(\frac{i}{2D}) - 2arctan(\frac{(i-s)}{2D})$

For uncrossed disparity $s > 0$. Resulting in $2arctan\frac{i-s}{2D} < 2arctan\frac{i}{2D}$.

Therefore $VD > 0$.

From the above it can be established that vergence difference is not the same as the HVA:

$$2arctan(\frac{i}{2D}) - 2arctan(\frac{(i-s)}{2D}) \neq 2arctan(\frac{s}{2D}) \tag{B.2}$$

Wartell [110] and Hodge [39] use vergence difference and HVA interchangeably and calculate the limits described by Yeh and Silverstein [121]; and Valyus [98] in such terms.

Appendix C

15" Display Calculations

Table C.1 shows the results of the limits applied to a 15" display with $D = 700mm$ and $i = 65mm$.

Siegel and Nagata [92] stated limits on microstereopsis of 1mm screen parallax being detected by 2/3rds of observers.			
Crossed: 3.35841, Uncrossed: 3.358409	Uncrossed: 1, Crossed: 1	Crossed: 10.606 Uncrossed: 10.94	Crossed:4.90034, Uncrossed: 4.900665
Akka [1] states that the limit on screen parallax should be a maximum of 3.5% of screen width			
Crossed: 35.8275, Uncrossed: 35.82751	Crossed: 10.67, Uncrossed: 10.668	Crossed: 98.689 Uncrossed: 137.44	Crossed: 52.2591, Uncrossed: 52.29602
Yeh and Silverstein [121] state that the limit should not exceed 27 minarc crossed and 24 minarc uncrossed disparity (for $< 200ms$)			
Crossed: 18.5071, Uncrossed: 16.44506	Crossed: 5.511, Uncrossed: 4.897	Crossed: 54.708, Uncrossed: 57.03	Crossed: 27, Uncrossed: 24
Yeh and Silverstein [121] state that the limit should not exceed 4.93° crossed and 1.53° uncrossed disparity (for $> 2s$)			
Crossed: 203.249, Uncrossed: 62.87707	Crossed: 60.52, Uncrossed: 18.722	Crossed: 337.506, Uncrossed: 283.19	Crossed: 295.8, Uncrossed: 91.8
Farrell and Booth [23] stated that the display to point of vergence limit should not exceed +/- 0.75D			
Crossed: 114.6279, Uncrossed: 114.618	Crossed: 34.132, Uncrossed: 34.13	Crossed: 241, Uncrossed: 774	Crossed: 167.025, Uncrossed: 167.4181
Valyus [98] stated that the point of focus and vergence difference should not exceed 1.6^o.			
Crossed: 65.8376,Uncrossed: 65.7523	Crossed: 19.578, Uncrossed: 19.6	Crossed: 162.199, Uncrossed: 301.73	Crossed: 96, Uncrossed: 96
Valyus [98] stated that screen parallax should be no more than 3% viewing distance.			
Crossed: 70.5266, Uncrossed: 70.5266	Crossed: 21, Uncrossed: 21	Crossed: 170.93, Uncrossed: 334.09	Crossed: 102.831, Uncrossed: 102.9744
Williams and Parrish [116] stated a limit of -25% - +60% for depth from screen plane.			
Crossed: 72.7655, Uncrossed: 81.86123	Crossed: 21.67, Uncrossed: 24.375	Crossed: 175, Uncrossed: 420	Crossed: 106.093, Uncrossed: 119.5341
Patterson and Martin [78] state that at the fovea, the maximum angle of disparity before diopia occurs is $1/10^o$.			
Crossed: 4.11208, Uncrossed: 4.11175	Crossed: 1.224, Uncrossed: 1.22	Crossed: 12.942, Uncrossed: 13.44	Crossed: 6, Uncrossed: 6
Jones et al. [48] stated that the projected disparity should not exceed 100mm behind or in front of the display screen			
Crossed: 36.3828, Uncrossed: 27.28708	Crossed: 10.83, Uncrossed: 8.125	Crossed: 100, Uncrossed: 100	Crossed: 53.0687, Uncrossed: 39.82685
Woods et al. [119] discovered that 75% of observers could fuse images where the screen parallax was -40mm - +45mm			
Crossed: 134.336, Uncrossed: 151.1284	Crossed: 40, Uncrossed: 45	Crossed: 266.667, Uncrossed: 1575	Crossed: 195.708, Uncrossed: 220.7756
Lipton [60] suggests that to reduce visual fatigue a $1/3rd$ of the limit described by Valyus'[98] should be applied.			
Crossed: 21.9351, Uncrossed: 21.92564	Crossed: 6.53, Uncrossed: 6.529	Crossed: 63.916, Uncrossed: 78.16	Crossed: 32, Uncrossed: 32

Table C.1: Limits on a 15" display

Appendix D

Derivation of Frame Cancellation Limits

The model shown in Figure 3.4 can be used to calculated the inequalities for the frame cancellation zones and derive the equations given in Table 3.1 and using Figures D.1 and D.2.

Firstly, the mathematics of the line equation is given by equation:

$y = mx + c$

Where m is the gradient of the line ($\frac{d(y)}{d(x)}$) and c the point of intersection when $x = 0$.

Let:
x be the horizontal displacement along the display where the left side of the display is taken as (0,0).
Sd be the distance from the display plane.
w be the width of the screen plane.
D be the distance from the observer to the screen.
i be the interocular distance.
c be the point of intersection when $x = 0$.

Dealing with the left inequalities initially and line 1 as shown in figure D.1:

$\frac{d(Sd)}{d(x)} = \frac{D}{\frac{w+i}{2}} = \frac{2D}{w+i}$

When $x = 0 \rightarrow Sd = 0$:

Inserting this into the line equation gives:

$Sd = \frac{2Dx}{w+i}$

Similarly for Line 2 in figure D.1:

$\frac{d(Sd)}{d(x)} = \frac{D}{\frac{w-i}{2}} = \frac{2D}{w-i}$

When $x = 0 \rightarrow Sd = 0$ giving:

$Sd = \frac{2Dx}{w-i}$

Now examining the right side of the model. Starting with Line 1 in figure D.2:

$\frac{d(Sd)}{d(x)} = -(\frac{D}{\frac{w+i}{2}}) = -(\frac{2D}{w+i})$

When $c = 0 \rightarrow Sd = ?$:

The value of c when $x = 0$ can be calculated from the ratio of the triangles in figure D.4 giving:

$c = D(\frac{w-i}{w+i}) + D$

Using these to construct the equation of the line:

$Sd = -\frac{2Dx}{w+i} + D\frac{w-i}{w+i} + D$

This can be simplified to:

$Sd = \frac{-2Dx+D(w-i)}{w+i} + D$

Similarly for Line 2 in figure D.2:

$\frac{d(Sd)}{d(x)} = -(\frac{D}{\frac{w-i}{2}}) = -(\frac{2D}{w-i})$

When $c = 0 \rightarrow Sd = ?$:

The value of c when $x = 0$ can be calculated from the ratio of the triangles in figure D.4 giving:

$c = D(\frac{w+i}{w-i}) + D$

Using these to construct the equation of the line:

$Sd = -\frac{2Dx}{w-i} + D\frac{w+i}{w-i} + D$

This can be simplified to:

$Sd = \frac{-2Dx+D(w+i)}{w-i} + D$

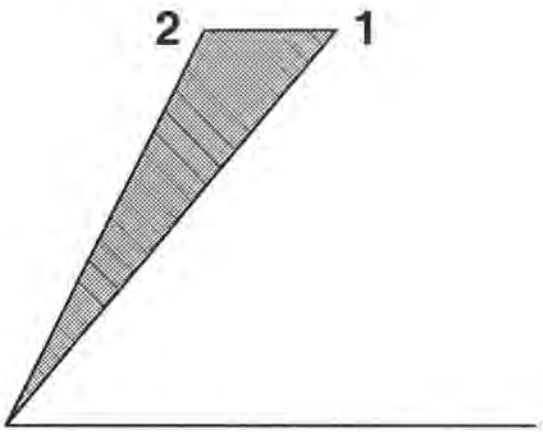


Figure D.1: Frame Cancellation - Left

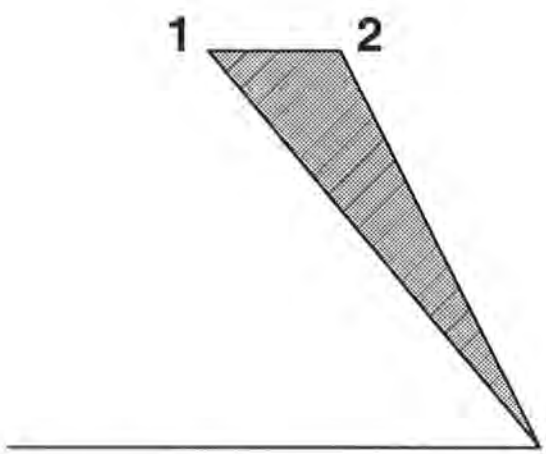


Figure D.2: Frame Cancellation - Right

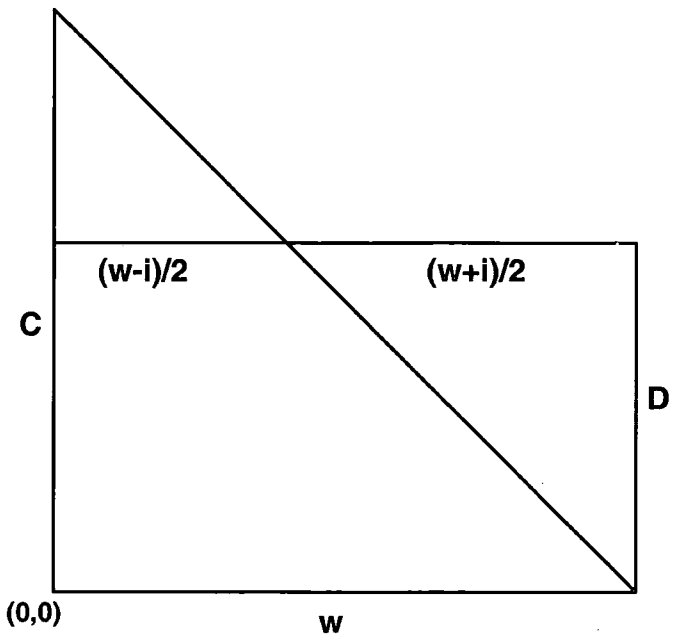


Figure D.3: Frame Cancellation - Right Line 1

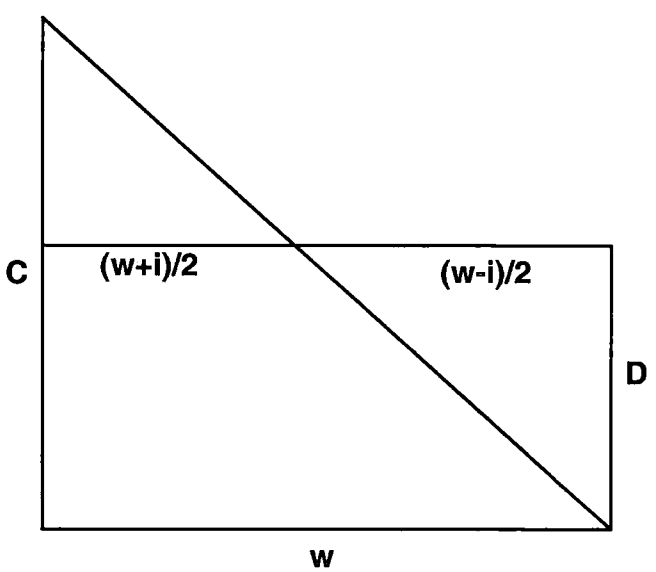


Figure D.4: Frame Cancellation - Right Line 2

Appendix E

The Test Images

Figure E.1: Middlebury Small - Barn1

Figure E.2: Middlebury Small - Barn2

Figure E.3: Middlebury Small - Bull

Figure E.4: Middlebury Small - Poster

Figure E.5: Middlebury Small - Sawtooth

Figure E.6: Middlebury Small - Venus

Figure E.7: Middlebury Large - Cones

Figure E.8: Middlebury Large - Teddy

Figure E.9: Sharp - Farm

Figure E.10: Sharp - Hand

Figure E.11: Scanned Image - As

Figure E.12: Downloaded Image - Flower

Appendix F

Disparity Analysis

F.1 Disparity Analysis - Single Images

	Total	Crossed	Uncrossed
Barn1	46.748	35.180	11.567
Barn2	21.239	15.111	6.127
Bull	20.239	13.476	6.763
Poster	54.996	30.998	23.997
Sawtooth	35.894	15.455	20.438
Venus	36.794	29.628	7.166

Table F.1: Small Middlebury (groundtruth) - S+N Limit Only

	Total	Crossed	Uncrossed
Small Cones			
Y+S	23.402	23.393	0.008
S+N	14.660	9.655	5.004
Lipton	11.455	11.454	0.001
Small Teddy			
Y+S	1.896	1.896	0
S+N	3.436	3.090	0.346
Lipton	1.123	1.123	0
Cones			
Y+S	43.476	36.905	6.570
S+N	7.651	6.551	1.099
Lipton	34.912	34.809	0.103
Teddy			
Y+S	41.683	15.101	26.582
S+N	0.396	0.396	0.000
Lipton	26.703	11.142	15.560

Table F.2: Large Middlebury (groundtruth)

	Total	Crossed	Uncrossed
Hand			
Y+S	43.578	36.089	7.488
S+N	4.537	2.772	1.764
Lipton	37.684	32.577	5.106
Farm			
Y+S	39.040	14.946	24.093
S+N	41.564	1.810	39.754
Lipton	26.035	12.982	13.053

Table F.3: Sharp Images

	Total	Crossed	Uncrossed
Flower	10.745	5.941	4.804
As	16.874	8.771	8.102

Table F.4: Miscellaneous Images - S+N Limit

F.2 Disparity Analysis - Segmented images

F.2.1 Small Middlebury - S+N Limit

	Average	Varience	Max	Min	Range	Median	Skew
Total	53.780	1371.022	100	0	100	52.068	-0.157
Crossed	35.275	1143.527	99.971	0	99.971	32.348	0.743
Uncrossed	18.504	937.380	100	0	100	0	1.620

Table F.5: Barn1 (groundtruth)

	Average	Varience	Max	Min	Range	Median	Skew
Total	20.279	1223.037	100	0	100	0	1.563
Crossed	15.092	1057.226	100	0	100	0	2.088
Uncrossed	5.187	209.507	60.89	0	60.89	0	3.432

Table F.6: Barn2 (groundtruth)

	Average	Varience	Max	Min	Range	Median	Skew
Total	18.727	640.411	65.777	0	65.777	0	0.866
Crossed	13.161	448.279	62.407	0	62.407	0	1.479
Uncrossed	5.565	152.745	45.79	0	45.79	0	2.439

Table F.7: Bull (groundtruth)

	Average	Varience	Max	Min	Range	Median	Skew
Total	51.467	1738.849	100	0	100	55.48	-0.091
Crossed	27.868	1357.270	99.891	0	99.891	9.820	1.085
Uncrossed	23.599	1313.080	100	0	100	0	1.293

Table F.8: Poster (groundtruth)

	Average	Varience	Max	Min	Range	Median	Skew
Total	30.963	1148.707	100	0	100	20.045	0.729
Crossed	13.921	439.252	56.423	0	56.423	0.965	1.281
Uncrossed	17.042	1092.860	100	0	100	0	1.893

Table F.9: Sawtooth (groundtruth)

	Average	Varience	Max	Min	Range	Median	Skew
Total	40.720	1633.105	100	0	100	30.147	0.424
Crossed	33.993	1550.692	100	0	100	18.19	0.792
Uncrossed	6.727	159.277	43.895	0	43.895	0	2.087

Table F.10: Venus (groundtruth)

F.2.2 Large Middlebury

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	31.143	1250.586	91.546	0.012	91.534	16.51	0.862
Y+S Crossed	27.091	1303.077	90.164	0	90.164	3.312	0.902
Y+S Uncrossed	4.052	46.129	22.86	0	22.86	0.995	1.871
S+N Total	14.045	421.877	61.07	0	61.07	1.42	1.286
S+N Crossed	9.873	235.922	53.02	0	53.02	0.615	1.609
S+N Uncrossed	4.171	73.499	30.43	0	30.43	0.005	2.193
Lipton Total	17.536	542.028	72.879	0.012	72.866	4.025	1.507
Lipton Crossed	13.489	565.304	72.879	0	72.879	0	1.791
Lipton Uncrossed	4.046	46.152	22.86	0	22.86	0.995	1.872

Table F.11: Small Cones (groundtruth)

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	1.051	2.865	6.797	0	6.797	0.022	2.295
Y+S Crossed	0.106	0.040	0.630	0	0.630	0	1.669
Y+S Uncrossed	0.944	2.367	6.166	0	6.166	0.022	2.300
S+N Total	2.934	37.089	21.194	0	21.194	0	2.180
S+N Crossed	2.91	36.181	20.72	0	20.72	0	2.151
S+N Uncrossed	0.024	0.011	0.474	0	0.474	0	4.471
Lipton Total	0.132	0.047	0.878	0	0.878	0.002	2.366
Lipton Crossed	0.009	0.000	0.078	0	0.078	0	2.417
Lipton Uncrossed	0.122	0.039	0.800	0	0.800	0.002	2.300

Table F.12: Small Teddy (groundtruth)

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	45.188	1792.319	100	0	100	24.13	0.360
Y+S Crossed	38.735	2030.452	100	0	100	8.73	0.462
Y+S Uncrossed	6.453	115.411	51.2	0	51.2	0.73	2.029
S+N Total	6.666	369.708	89.9	0	89.9	0	3.368
S+N Crossed	5.806	292.828	79.64	0	79.64	0	3.320
S+N Uncrossed	0.859	10.050	20.76	0	20.76	0	5.210
Lipton Total	41.108	1791.807	100	0	100	22.79	0.494
Lipton Crossed	36.534	1918.694	100	0	100	3.635	0.548
Lipton Uncrossed	4.574	62.084	28.18	0	28.18	0.595	1.875

Table F.13: Cones (groundtruth)

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	51.986	1557.748	100	0	100	64.885	-0.130
Y+S Crossed	20.193	1118.765	99.82	0	99.82	0	1.271
Y+S Uncrossed	31.793	1449.811	100	0	100	9.855	0.879
S+N Total	0.432	4.684	14.16	0	14.16	0	5.658
S+N Crossed	0.432	4.684	14.16	0	14.16	0	5.658
S+N Uncrossed	0.000	0	0.02	0	0.02	0	5.933
Lipton Total	35.209	1354.954	100	0	100	13.315	0.442
Lipton Crossed	15.435	733.833	84.12	0	84.12	0	1.425
Lipton Uncrossed	19.774	1022.683	100	0	100	3.085	1.588

Table F.14: Teddy (groundtruth)

F.2.3 Sharp Images

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	45.482	985.803	99.99	0	99.99	53.015	-0.218
Y+S Crossed	15.969	573.288	81	0	81	1.841	1.549
Y+S Uncrossed	29.512	899.315	99.99	0	99.99	12.712	0.674
S+N Total	32.595	1192.729	100	0	100	21.785	0.936
S+N Crossed	1.763	7.915	16.33	0	16.33	0.93	3.683
S+N Uncrossed	30.832	1236.541	100	0	100	19.54	0.966
Lipton Total	33.824	994.958	99.7	0	99.7	25.075	0.396
Lipton Crossed	14.151	519.985	77.840	0	77.840	1.405	1.685
Lipton Uncrossed	19.673	704.592	99.7	0	99 7	6.846	1.494

Table F.15: Farm

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	42.418	564.463	88.032	1.559	86.473	46.919	-0.103
Y+S Crossed	22.570	534.698	84.987	0	84.987	14.990	1.366
Y+S Uncrossed	19.848	372.699	75	0	75	12.116	0.882
S+N Total	5.575	17.896	15.533	0.229	15.304	3.921	0.842
S+N Crossed	3.181	8.060	13.988	0	13.988	2.441	1.956
S+N Uncrossed	2.393	4.516	8.871	0	8.871	1.698	1.384
Lipton Total	34.989	525.682	85.549	1.126	84.422	36.839	0.191
Lipton Crossed	18.693	514.520	83.096	0	83.096	8.970	1.474
Lipton Uncrossed	16.296	284.655	64.123	0	64.123	8.480	0.988

Table F.16: Hand

F.2.4 Miscellaneous Images

	Average	Varience	Max	Min	Range	Median	Skew
Total	14.394	38.269	29.375	3.94	25.435	13.24	0.834
Crossed	7.030	5.492	12.602	2.49	10.112	7.29	0.070
Uncrossed	7.364	25.840	21.461	1.45	20.011	5.703	1.672

Table F.17: As - S+N Limit

	Average	Varience	Max	Min	Range	Median	Skew
Total	7.962	15.639	15.418	2.231	13.186	7.165	0.731
Crossed	4.327	6.249	9.449	0.787	8.661	4.177	0.973
Uncrossed	3.635	2.569	5.969	1.444	4.525	3.301	0.327

Table F.18: Flower - S+N Limit

F.3 Disparity Analysis - Weighted images

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	31.14	1250.59	91.55	0.012	91.53	16.51	0.86
Y+S Crossed	27.09	1303.08	90.16	0	90.16	3.31	0.90
Y+SUncrossed	4.05	46.13	22.86	0	22.86	1.00	1.87
Lipton Total	17.54	542.03	72.88	0.01	72.87	4.03	1.51
Lipton Crossed	13.49	565.30	72.88	0	72.88	0	1.79
Lipton Uncrossed	4.05	46.15	22.86	0	22.86	1.00	1.87

Table F.19: SmallCones (groundtruth)

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	2.941	11.639	11.273	0.006	11.267	1.407	1.227
Y+S Crossed	0.945	3.362	6.213	0	6.213	0.008	2.258
Y+S Uncrossed	1.995	11.223	11.273	0	11.273	0.490	1.873
Lipton Total	0.379	0.262	1.750	0.000	1.749	0.116	1.580
Lipton Crossed	0.069	0.035	0.717	0	0.717	0	2.943
Lipton Uncrossed	0.309	0.270	1.750	0	1.750	0.076	1.873

Table F.20: Small Teddy (groundtruth)

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	34.981	1560.880	100	0	100	14.083	0.717
Y+S Crossed	31.352	1631.508	100	0	100	1.437	0.761
Y+S Uncrossed	3.629	49.150	28.18	0	28.18	0.323	2.242
Lipton Total	26.276	1025.161	100	0	100	12.634	1.139
Lipton Crossed	22.774	1069.079	100	0	100	0.180	1.249
Lipton Uncrossed	3.502	48.795	28.18	0	28.18	0.255	2.265

Table F.21: Cones (groundtruth)

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	26.799	717.791	98.271	0	98.271	11.982	0.707
Y+S Crossed	9.563	289.738	68.082	0	68.082	0.738	2.027
Y+S Uncrossed	17.236	592.042	98.271	0	98.271	5.998	1.685
Lipton Total	19.257	466.626	87.752	0	87.752	6.998	1.204
Lipton Crossed	6.848	189.221	64.018	0	64.018	0.382	2.693
Lipton Uncrossed	12.409	368.765	87.752	0	87.752	3.187	2.156

Table F.22: Sharp Farm

F.4 StereoMatch Disparity Maps

F.4.1 Single Images

	Total	Crossed	Uncrossed
Barn1	16.963	5.201	11.761
Barn2	8.488	4.575	3.913
Bull	18.787	6.440	12.347
Poster	50.618	24.248	26.369
Sawtooth	38.694	16.526	22.168
Venus	22.885	6.244	16.640

Table F.23: Small Middlebury - S+N limit

	Total	Crossed	Uncrossed
Small Cones			
Y+S	10.763	6.367	4.396
S+N	17.788	10.521	7.267
Lipton	4.779	1.788	2.991
Small Teddy			
Y+S	15.216	1.630	13.585
S+N	31.323	21.640	9.682
Lipton	6.553	1.210	5.342
Cones			
Y+S	63.082	29.048	34.033
S+N	9.722	5.511	4.210
Lipton	56.104	27.845	28.259
Teddy			
Y+S	55.591	7.694	47.896
S+N	19.852	13.530	6.322
Lipton	47.728	5.745	41.982

Table F.24: Large Middlebury

F.4.2 Small Middlebury - Segmented Images

	Average	Varience	Max	Min	Range	Median	Skew
Total	18.740	715.237	81.639	0	81.639	1.598	1.309
Crossed	6.952	273.993	57.289	0	57.289	0.190	2.585
Uncrossed	11.787	400.819	76.284	0	76.284	1.191	2.244

Table F.25: Barn1

	Average	Varience	Max	Min	Range	Median	Skew
Total	9.565	197.132	45.948	0	45.948	2.912	1.681
Crossed	5.691	121.518	42.068	0	42.068	0.725	2.374
Uncrossed	3.874	78.952	40.271	0	40.271	1.148	3.985

Table F.26: Barn2

	Average	Varience	Max	Min	Range	Median	Skew
Total	18.933	1121.408	97.827	0	97.827	0.770	1.627
Crossed	6.464	220.985	51.777	0	51.777	0.371	2.488
Uncrossed	12.468	458.367	60.395	0	60.395	0.380	1.634

Table F.27: Bull

	Average	Varience	Max	Min	Range	Median	Skew
Total	53.642	1520.485	99.713	1.451	98.261	61.880	-0.179
Crossed	30.348	1472.679	99.513	0.040	99.473	9.298	0.971
Uncrossed	23.293	928.406	93.753	0	93.753	5.627	1.175

Table F.28: Poster

	Average	Varience	Max	Min	Range	Median	Skew
Total	40.955	1834.435	100	0	100	32.344	0.183
Crossed	19.691	1161.407	88.933	0	88.933	0.460	1.295
Uncrossed	21.264	830.694	75.49	0	75.49	4.597	1.060

Table F.29: Sawtooth

	Average	Varience	Max	Min	Range	Median	Skew
Total	21.531	701.850	85.933	0.17	85.763	9.881	1.332
Crossed	6.078	232.060	63.822	0.04	63.782	1.057	3.412
Uncrossed	15.452	342.838	54.592	0	54.592	6.105	0.996

Table F.30: Venus

F.4.3 Large Middlebury - Segmented

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	11.844	240.192	51.732	0.110	51.621	3.786	1.564
Y+S Crossed	5.326	129.808	46.261	0	46.261	0.513	2.904
Y+S Uncrossed	6.517	113.168	34.061	0	34.061	0.754	1.625
S+N Total	16.505	440.935	76.579	0	76.579	5.601	1.591
S+N Crossed	9.077	215.524	60.411	0	60.411	2.754	2.539
S+N Uncrossed	7.427	92.627	27.380	0	27.380	2.267	1.100
Lipton Total	5.605	85.733	31.203	0.080	31.123	1.222	1.883
Lipton Crossed	0.559	0.768	3.616	0	3.616	0.222	2.671
Lipton Uncrossed	5.045	85.938	30.233	0	30.233	0.310	1.870

Table F.31: Small Cones

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	17.334	461.998	79.18	0.441	78.738	7.411	1.924
Y+S Crossed	1.070	2.274	5.369	0	5.369	0.356	1.643
Y+S Uncrossed	16.263	485.085	79.18	0.400	78.779	6.346	1.888
S+N Total	30.723	666.840	77.960	0	77.960	32.917	0.185
S+N Crossed	21.341	366.127	64.563	0	64.563	20.795	0.512
S+N Uncrossed	9.382	143.959	39.275	0	39.275	4.141	1.385
Lipton Total	9.166	143.200	49.24	0.21	49.03	3.689	2.236
Lipton Crossed	0.544	0.465	1.959	0	1.959	0.135	1.002
Litpon Uncrossed	8.621	148.277	49.24	0.21	49.03	3.216	2.217

Table F.32: Small Teddy

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	65.050	1143.637	100	0.34	99.66	79.057	-0.740
Y+S Crossed	31.076	1532.955	99.298	0	99.298	4.294	0.779
Y+S Uncrossed	33.974	1503.953	100	0	100	6.714	0.609
S+N Total	9.132	350.062	93.317	0	93.317	1.467	2.930
S+N Crossed	5.170	179.663	68.12	0	68.12	0.670	3.822
S+N Uncrossed	3.961	79.735	47.486	0	47.486	0.564	3.119
Lipton Total	58.376	1179.507	99.98	0.34	99.64	67.288	-0.467
Lipton Crossed	29.834	1471.275	99.136	0	99.136	3.092	0.809
Lipton Uncrossed	28.541	1261.758	99.98	0	99.98	4.809	0.840

Table F.33: Cones

	Average	Varience	Max	Min	Range	Median	Skew
Y+S Total	55.554	1158.417	99.89	0	99.89	54.408	-0.146
Y+S Crossed	8.700	224.657	67.699	0	67.699	1.045	2.010
Y+S Uncrossed	46.854	1569.497	99.86	0	99.86	48.416	0.120
S+N Total	19.079	650.318	97.29	0	97.29	7.63	1.642
S+N Crossed	12.998	359.024	79.067	0	79.067	4.154	1.867
S+N Uncrossed	6.081	90.156	40.598	0	40.598	1.975	2.190
Lipton Total	47.802	1148.237	99.85	0	99.85	42.936	0.172
Lipton Crossed	6.509	122.670	41.415	0	41.415	0.866	1.894
Lipton Uncrossed	41.293	1430.891	99.74	0	99.74	29.172	0.332

Table F.34: Teddy

Appendix G

Intensity Level Analysis

G.1 Intensity Level Analysis of Stereo Images

G.1.1 Small Middlebury - Single

	Average	Red	Green	Blue
Sawtooth	2.643	2.541	2.352	3.036
Bull	1.547	1.595	1.26	1.786
Poster	2.969	2.953	2.727	3.227
Barn1	2.223	2.448	1.947	2.275
Barn2	2.092	1.948	1.606	2.222
Venus	1.925	2.05	1.704	2.021

Table G.1: Small Middlebury - Groundtruth

G.1.2 Large Middlebury - Single

	Average	Red	Green	Blue
Small Teddy	3.071	2.733	2.902	3.577
Teddy	3.026	2.716	2.88	3.483
Small Cones	3.72	3.745	3.73	3.684
Cones	3.612	3.621	3.627	3.589

Table G.2: Large Middlebury - Groundtruth

G.1.3 Sharp Images - Single

	Average	Red	Green	Blue
Sharp Farm	1.787	1.929	1.719	1.715
Sharp Hand	1.691	1.51	1.717	1.846

Table G.3: Sharp Images - StereoMatch

G.1.4 Miscellaneous Images - Single

	Average	Red	Green	Blue
As	3.174	3.009	3.426	3.087
Flower	2.265	2.327	2.226	2.242

Table G.4: Miscellaneous Images - StereoMatch

G.2 Intensity Difference Analysis of Segmented Images

G.2.1 Small Middlebury

	Average	Varience	Max	Min	Range	Median	Skew
Average	2.195	0.535	3.381	1.096	2.285	2.138	0.088
Red	2.421	0.805	3.810	1.083	2.727	2.333	0.137
Green	1.915	0.731	3.310	0.677	2.633	1.886	0.155
Blue	2.249	0.232	3.076	1.443	1.633	2.221	0.063

Table G.5: Barn1 - Groundtruth

	Average	Varience	Max	Min	Range	Median	Skew
Average	1.912	1.004	4.724	1.263	3.461	1 544	2.325
Red	1.881	0.850	4.801	1.237	3.563	1.517	2.611
Green	1.584	0.654	3.633	0.980	2.652	1.266	2.049
Blue	2.271	1.718	5.740	1.537	4.202	1.761	2.341

Table G.6: Barn2 - Groundtruth

	Average	Varience	Max	Min	Range	Median	Skew
Average	1.603	0.093	2.134	1.137	0.996	1.619	-0.005
Red	1.268	0.187	2.322	0.706	1.616	1.177	0.694
Green	1.785	0.062	2.192	1.397	0.795	1.789	0.026
Blue	1.552	0.093	2.132	1.080	1.051	1.561	-0.007

Table G.7: Bull - Groundtruth

	Average	Varience	Max	Min	Range	Median	Skew
Average	2.875	1.047	5.522	1.775	3.747	2.584	1.316
Red	2.659	1.074	5.306	1.202	4.103	2.437	1.486
Green	3.182	0.615	5.164	2.192	2.972	2.940	1.358
Blue	2.905	0.858	5.331	1.754	3.577	2.728	1.472

Table G.8: Poster - Groundtruth

	Average	Varience	Max	Min	Range	Median	Skew
Average	2.711	3.663	9.694	0.998	8.695	2.238	2.690
Red	2.596	4.658	10.411	0.648	9.763	2.022	2.708
Green	3.244	2.946	9.312	1.707	7.604	2.643	2.578
Blue	2.850	3.692	9.806	1.131	8.674	2.285	2.676

Table G.9: Sawtooth - Groundtruth

Average	Varience	Max	Min	Range	Median	Skew
2.025	0.297	2.861	0.962	1.898	1.957	-0.229
2.185	0.396	3.340	1.048	2.291	2.127	-0.046
1.824	0.372	2.756	0.668	2.087	1.760	-0.242
2.067	0.184	2.810	1.169	1.641	2.067	-0.158

Table G.10: Venus - Groundtruth

G.2.2 Large Middlebury

	Average	Varience	Max	Min	Range	Median	Skew
Average	3.624	1.692	6.514	1.965	4.549	3.175	1.266
Red	3.797	2.083	6.581	1.752	4.829	3.484	0.925
Green	3.596	2.229	7.088	1.850	5.238	3.029	1.287
Blue	3.479	1.280	6.710	2.293	4.417	3.116	1.884

Table G.11: Small Cones - Groundtruth

	Average	Varience	Max	Min	Range	Median	Skew
Average	3.160	1.865	6.642	1.307	5.334	2 927	1.545
Red	2.809	1.330	5.515	1.254	4.260	2.707	1.318
Green	3.050	2.624	7.179	0.976	6.203	2.928	1.327
Blue	3.621	2.010	7.550	1.689	5.860	3.243	1.828

Table G.12: Small Teddy - Groundtruth

	Average	Varience	Max	Min	Range	Median	Skew
Average	3.495	2.457	8.438	0	8.438	3.025	1.206
Red	3.461	2.726	8.400	0	8.400	3.116	1.186
Green	3.511	3.715	9.765	0	9.765	2.825	1.512
Blue	3.512	2.191	8.836	0	8.836	2.945	1.489

Table G.13: Cones - Groundtruth

	Average	Varience	Max	Min	Range	Median	Skew
Average	2.991	4.241	13.211	1.318	11.893	2.540	3.447
Red	2.712	2.843	12.255	1.312	10.943	2.329	3.628
Green	2.817	5.976	13.836	1.023	12.812	2.174	3.158
Blue	3.444	4.717	13.880	1.510	12.370	2.995	3.587

Table G.14: Teddy - Groundtruth

G.2.3 Sharp Images

	Average	Varience	Max	Min	Range	Median	Skew
Average	6.076	8.804	12.634	1.194	11.439	5.390	0.716
Red	6.449	10.488	14.470	1.519	12.951	5.326	0.774
Green	7.631	14.687	15.528	1.391	14.136	7.370	0.567
Blue	6.719	10.248	12.709	1.368	11.341	5.789	0.544

Table G.15: Sharp Hand

	Average	Varience	Max	Min	Range	Median	Skew
Average	6.366	17.708	17.487	0.998	16.488	4.743	1.205
Red	7.710	31.457	20.951	1.111	19.840	5.815	1.010
Green	6.895	20.916	20.386	1.130	19.255	5.505	1.589
Blue	4.494	9.470	14.843	0.752	14.091	3.340	1.810

Table G.16: Sharp Farm

G.2.4 Miscellaneous

	Average	Varience	Max	Min	Range	Median	Skew
Average	4.155	3.643	7.032	1.107	5.924	4.230	-0.122
Red	4.721	4.545	8.211	1.342	6.868	4.860	-0.071
Green	4.120	2.727	7.025	1.194	5.830	3.956	-0.088
Blue	4.332	3.521	7.112	1.215	5.897	4.432	-0.124

Table G.17: As

	Average	Varience	Max	Min	Range	Median	Skew
Average	3.786	13.741	11.926	0.582	11.343	2.192	1.563
Red	3.687	13.212	11.872	0.703	11.168	2.167	1.733
Green	3.783	12.986	12.029	0.849	11.180	2.134	1.870
Blue	3.752	13.243	11.943	0.761	11.181	2.241	1.737

Table G.18: Flower

G.3 Intensity Difference - StereoMatch

G.3.1 Single Images

	Average	Red	Green	Blue
SmallCones	3.228	3.315	3.085	3.284
Cones	2.836	2.879	2.705	2.924
SmallTeddy	2.664	2.654	2.216	3.123
Teddy	2.434	2.426	2.059	2.818

Table G.19: Large Middlebury

G.3.2 Segmented Images

	Average	Varience	Max	Min	Range	Median	Skew
Average	3.334	1.142	6.140	1.919	4.221	2.970	1.212
Red	3.435	1.284	6.246	1.699	4.546	3.179	0.834
Green	3.195	1.660	7.441	1.811	5.629	2.830	2.087
Blue	3.373	0.890	5.548	2.245	3.303	3.031	1.075

Table G.20: SmallCones

	Average	Varience	Max	Min	Range	Median	Skew
Average	2.721	1.467	6.746	1.286	5.460	2.391	2.108
Red	2.686	1.970	6.577	1.210	5.366	2.320	2.020
Green	2.267	1.326	6.398	0.945	5.452	2.098	2.501
Blue	3.211	1.491	7.264	1.673	5.591	2.841	2.050

Table G.21: SmallTeddy

	Average	Varience	Max	Min	Range	Median	Skew
Average	2.970	4.344	15.418	1.641	13.776	2.395	4.125
Red	3.035	5.441	15.043	1.321	13.722	2.583	3.564
Green	2.850	8.036	23.833	1.233	22.599	2.088	6.089
Blue	3.026	1.964	10.276	1.981	8.295	2.596	3.315

Table G.22: Cones

	Average	Varience	Max	Min	Range	Median	Skew
Average	2.593	5.005	12.693	1.217	11.475	1.952	3.322
Red	2.587	6.356	13.350	1.179	12.170	1.901	3.345
Green	2.209	4.817	13.180	0.939	12.241	1.551	3.731
Blue	2.984	4.516	12.399	1.381	11.017	2.483	3.191

Table G.23: Teddy

G.4 JEPG Images - Single Images

	Average	Red	Green	Blue
Venus	2.124	2.185	1.982	2.205
Bull	1.665	1.623	1.452	1.922
Barn1	2.627	2.742	2.437	2.703
Barn2	2.095	2.042	1.917	2.328
Poster	3.454	3.541	3.255	3.566
Sawtooth	2.740	2.734	2.589	2.898

Table G.24: Small Middlebury - Groundtruth

	Average	Red	Green	Blue
SmallCones	3.94	4.089	4.027	3.704
Cones	3.609	3.736	3.73	3.362
SmallTeddy	3.161	3.03	3.125	3.33
Teddy	2.955	2.756	2.934	3.175

Table G.25: Large Middlebury - Groundtruth

Appendix H

Frame Cancellation Results

	Left	Right
Barn1	53.595	15.682
Barn2	45.433	0
Bull	32.559	0.170
Poster	23.955	69.099
Sawtooth	48.842	37.578
Venus	50.861	58.642

Table H.1: Small Middlebury (groundtruth)

	Left	Right
Small Cones	46.706	51.066
Small Teddy	79.373	64.72
Cones	47.773	50.273
Teddy	83.126	65.386

Table H.2: Large Middlebury (groundtruth)

	Left	Right
Farm	0.204	81.994
Hand	82.188	47.461

Table H.3: Sharp

	Left	Right
As	23.847	33.065
Flower	5.782	7.064

Table H.4: Other Images

Bibliography

[1] R. Akka. Automatic software control of display parameters for stereoscopic graphic images. In *Stereoscopic Displays and Applications III*, volume 1669, pages 31–38 (June 1992).

[2] David Alais and Randolph Blake. *Binocular Rivalry.* Bradford Book (February 2005). ISBN 026201212X.

[3] David Alais, Robert P O'Shea, Corinne Mesana-Alias, and Ian G Wilson. On binocular alternation. *Perception*, 29:1437–1445 (2000).

[4] Tim Andrews. Binocular rivalry - luminance limits (Sept 2004). (Personal Communication).

[5] Timothy J. Andrews and D. Purves. Simularities in normal and binocular rivalrous viewing. In *Proc Natl Acad Sci U S A.*, volume 94, pages 9905–9908 (Sept 1997).

[6] G. Balakrishnan, Fitzmaurice and G. Kurtenbach. Userinterfaces for volumetric displays. *IEEE Computing*, pages 37–45 (March 2001).

[7] Guy E. Blelloch. Introduction to data compression (October 2001). Accessed Jan 2005, URL `http://www-2.cs.cmu.edu/afs/cs/project/pscico-guyb/realworld/www/compression.pdf`.

[8] M. Bleyer and M. Gelautz. A layered stereo algorithm using image segmentation and global visibility constraints. *ICIP*, 5:2997–3000 (Oct 2004).

[9] P. Burt and B. Julesz. Modifications of the classical notion of panum's fusional area. *Perception*, 9(6):671–682. (1980).

[10] Thomas A. Carlson and Sheng He. Competing global representations fail to initiate binocular rivalry. *Neuron*, 43:97–914 (Sept 2004).

[11] J. A. Cegalis. Prism distortion and accommodative change. *Perception and Psychophysics*, 13(3):494–498 (1973).

[12] Alan Chalmers, Ann McNamara, Scott Daly, Karol Myszkowski, and Tom Troscianko. *Image Quality Metrics.* ACM SIGGRAPH (July 2000).

[13] Zezhi Chen, Chengke Wu, and Hung Tat Tsui. A new image rectification algorithm. *Pattern Recogn. Lett.*, 24(1-3):251–260 (2003). ISSN 0167-8655. URL `http://dx.doi.org/10.1016/S0167-8655(02)00239-8`.

[14] R. Cooper (1998). Accessed May 2003, URL `http://www.vision3d.com/`.

[15] NVIDIA Corporation. Forceware graphics drivers: Nvidia 3d stereo user's guide (July 2004). Accessed Jan 2005, URL `http://download.nvidia.com/Windows/61.76/61.76_ForceWare_3D_Stereo_Users_Guide.pdf`.

[16] David Crascic and Paul Milgram. Perceptual issues in augmented reality. In *SPIE: Stereoscopic Displays and Virtual Reality Systems III*, volume 2653, pages 123–134 (Feb 1996).

[17] David Daniel. Institute of ophthalmology - eye disease. Accessed May 2003, URL `http://www.ucl.ac.uk/ioo/eyedisease/index.htm`.

[18] D.M.Szaflarski. How we see: The first steps of human vision, access excellence classic collection. Accessed December 2003, URL `http://www.accessexcellence.org/AE/AEC/CC/vision_background.html`.

[19] N. Dodgson, J. R. Moore, and S. R. Lang. Multi-view autostereoscopic 3d display. *International Broadcasting Convention*, pages 497–502 (Sept 1999).

[20] Neil A. Dodgson. Variation and extrema of human interpupillary distance. In *SPIE - IS&T Electronic Imaging XV*, volume 5291 (2004). ISSN 0277-786X.

[21] G. Egnal, M. Mintz, and R. Wildes. A stereo confidence metric using single view imagery. volume 162, pages 162–170. IV02 (2002).

[22] David Ezra, Graham J. Woodgate, Basil A. Omar, Nicolas S. Holliman, Jonathan Harrold, and Larry S. Shapiro. New autostereoscopic display system. In *Stereoscopic Displays and Virtual Reality Systems II*, volume 2409 (Feb 1995).

[23] R.J. Farrell and J.M. Booth. *Design Handbook for Imagery Interpretation Equipment.* Seattle, WA: Boeing Aerospace Co. (1984). Ref: D180-19063-1.

[24] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint.* The MIT Press, Cambridge (1993). ISBN 0-262-06158-9.

[25] J.A. Ferwerda. Perceiving size and space (2000). Accessed July 2003, URL `http://www.graphics.cornell.edu/~jaf/projects/pn/space.html`.

[26] W.J. Freeman. The physiology of perception. *Scientific American*, 264(2):78–85 (1991).

[27] Andrea Fusiello, Emanuele Trucco, and Alessandro Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22 (2000). URL `citeseer.ist.psu.edu/fusiello00compact.html`.

[28] Davi Geiger, Bruce Ladendorf, and Alan L. Yuille. Occlusions and binocular stereo. In *European Conference on Computer Vision*, pages 425–433 (1992).

[29] E. Bruce Goldstein. *Sensation and Perception.* Brooks Cole, fifth edition (Nov 1998). ISBN 0534346804.

[30] L. Gooding, M. Miller, J. Moore, and S. Kim. The effect of viewing distance and disparity on perceived depth. In *Stereoscopic Displays and Applications II*, volume 1457, pages 259–266. SPIE (Aug 1991).

[31] Daniel Gottlieb. Vision, gottlieb vision group. Accessed March 2003, URL `http://www.gottliebvisiongroup.com/dictionary.htm`.

[32] Richard L. Gregory. *Eye and Brain - The Psychology of Seeing.* Oxford University Press, fifth edition (1998). ISBN 0-19-852412-9.

[33] Michael Halle. Autostereoscopic displays and computer graphics. *Computer Graphics, ACM SIGGRAPH*, 2(31):58–62. (May 1997).

[34] J. M. Harris and A. J. Parker. Efficiency of stereopsis in random-dot stereograms. *Optical Society of America Journal A*, 9:14–24 (January 1992).

[35] Julie M. Harris and A.J.Parker. Constraints on human stereo dot matching. *Vision Res*, 34(20):2761–72 (Oct 1994).

[36] Julie M. Harris and A.J.Parker. Independant neural mechanisms for brightness and dark information in binocular stereopsis. *Nature*, 374(6525):808–811 (1995).

[37] D.J. Heeger. Linking visual perception with human brain activity. *Neurobiology*, 9(4):474–479 (1999).

[38] H.v. Helmholtz. *Treatise on Physiological Optics (translated by J.P.C. Southall)*, volume 1-2. Dover Publications (1962).

[39] Larry F. Hodges. Tutorial: Time-multiplexed stereoscopic computer graphics. *IEEE Computer Graphics and Applications*, 12(2):20–30 (March 1992).

[40] L.F. Hodges and E.T. Davis. Geometric considerations for stereoscopic virtual environments. *Presence*, 2(1):34–43 (1993).

[41] Nicolas Steven Holliman. 3d display systems. Accessed 2003, URL `http://citeseer.ist.psu.edu/591367.html`.

[42] N.S. Holliman. Frame cancellation phenomenon (Sept 2004). [Personal Communication].

[43] Ian P. Howard and Brian J. Rogers. *Binocular Vision and Stereopsis.* Oxford University Press (1995). ISBN 0-19-508476-4.

[44] i Art Corporation. i-magic plug-in - users manual (November 2004). Accessed January 2005, URL `http://www.iart3d.com/ENG/Downloads/i-MagicCameraPluginUser'sManual-ENG.pdf`.

[45] W.A. IJsselsteijn, P.J.H. Seuntiens, and L.M.J. Meesters. State-of-the-art in human factors and quality issues of stereoscopic broadcast television. Technical report, Eindhoven University of Technology, Department Technology Management (August 2002).

[46] J.W. Jainta S. Fixation disparity: binocular vergence accuracy for a visual display at different positions relative to the eyes. *Human Factors*, 44(3):443–450 (2002).

[47] Joseph S Lappin Jan J Koenderink, Chris Christou. Shape constancy in pictorial relief. *Perception*, 25(2):155–164 (1996).

[48] G. R. Jones, D. Lee, N. S. Holliman, and D. Ezra. Controlling perceived depth in stereoscopic images. In *Stereoscopic Displays and Virtual Reality Systems VIII, San Jose, California, SPIE*, volume 4297A (January 2001).

[49] Graham Jones. Sharp images - differences in intensity level (Sept 2004). (Personal Communication).

[50] Peter K. Kaiser. The joy of visual perception: A web book. Accessed November 2002, URL `http://www.yorku.ca/eye/`.

[51] Sing Bing Kang, R.Szeliski, and P Anandan. The geometry-image representation trade off for rendering. *ICIP*, 2:13–16 (Sept 2000).

[52] Yoshihiro Kawai, Toshio Ueshiba, Yutako Ishiyawa, Yasushi Sumi, and Fumraki Tomita. Stereo correspondance using segment connectivity.

In *ICPR Proceedings of the 14th International Conference on Pattern Recognition*, volume 1. IEEE Computer Society (1998).

[53] M. Kersenbrock. Pentax stereo adapter for slrs (1999). Accessed May 2003, URL `http://www.photo-3d.com/CamPages/pentax.html`.

[54] Gerhard Klimeck, Myche McAuley, Tom Cwik, Bob Deen, and Eric DeJong. Parallel algorithms for near-realtime visualization - algorithm for quality measurement of stereo image correlation. Accessed May 2004, URL `http://dynamo.ecn.purdue.edu/~gekco/mars/correlator_qual.html`.

[55] H. Kohno. The simple making of stereoscopic photography (2001). Accessed November 2003, URL `http://www.rpm.or.jp/home/h-kouno/3dphoto.htm`.

[56] Helga Kolb, Eduardo Fernandez, and Ralph Nelson. Webvision: The organisation of the retina and visual system. Accessed November 2002, URL `http://webvision.med.utah.edu/index.html`.

[57] U.M. Leloglu. Artificial verses natural stereo depth perception (1994). Accessed Nov 2004, URL `http://heaven.eee.metu.edu.tr/~vision/books/alife/ch6.html`.

[58] W. Levelt. *On binocular rivalry.* Psychological Studies (1968).

[59] Brian Lingard. A tutorial on time-multiplexed stereoscopic computer graphics (talk) (1995). Accessed December 2004, URL `http://www.cs.wpi.edu/~matt/courses/cs563/talks/stereohtml/stereo.html`.

[60] L. Lipton. *Foundations of the Stereoscopic Cinema: A Study in Depth.* Van Nostrand, New York (1982). ISBN 0-442-24724-9.

[61] Lenny Lipton. Stereo3d handbook: A guide to creating stereoscopic images for stereographics stereo3d displays. Accessed March 2003, URL `http://www.stereographics.com/support/downloads_support/handbook.pdf`.

[62] Alfred Lit. The magnitude of the pulfrich stereophenomenon as a function of binocular difference of intensity at various levels of illumination. *The American Journal of Psychology*, 62(2) (April 1949).

[63] Alexander D. Logvinenko, Julie Epelboim, and Robert M. Steinman. The role of vergence in the perception of distance: a fair test of bishop berkley's claim. *Spatial Vision*, 15:77–97 (2001).

[64] Qian N. Matthews N, Xu P. A physiological theory of depth perception from vertical disparity. *Vision Res*, 43(1):85–99 (2003).

[65] D.F. McAllister. *3D Displays, Encyclopaedia on Imaging and Technology.* John Wiley and Sons (Feb. 2002). ISBN 0471332992.

[66] J. S. McVeigh, Mel Siegel, and Angel Jordan. Algorithm for automated eye strain reduction in real stereoscopic images and sequences. In *Human Vision and Electronic Imaging*, volume 2657, pages 307 – 316 (February 1996).

[67] David J. Montgomery, Christopher K. Jones, James N. Stewart, and Alan Smith. Stereoscopic camera design. In *Proceedings of SPIE, Stereoscopic Displays and Virtual Reality Systems IX*, volume 4660 (2002).

[68] George Themelis & Jim Motley. A crash course in stereo photography. Accessed June 2004, URL `http://home.att.net/~osps/tutorial/index.htm`.

[69] R. Nave. The eye. Accessed January 2004, URL `http://hyperphysics.phy-astr.gsu.edu/hbase/hframe.html`.

[70] Sharp Research Lab. of Europe. Stereo images (2003). [Graphical Images].

[71] University of Manchester. Tina - open source image analysis environment. Accessed November 2003 [Computer Program], URL `http://www.tina-vision.net/`.

[72] University of Michigan Kellogg Eye Center. Eye conditions, diseases and disorders. Accessed November 2003, URL `http://www.kellogg.umich.edu/patientcare/conditions/index.html`.

[73] G. Olague, F. Fernandez, C. Perez, , and E. Lutton. The infection algorithm: An artificial epidemic approach for dense stereo correspondence. *Artificial Life, MIT Press* (2005).

[74] World Health Organisation. Magnitude and causes of visual impairment (November 2004). Accessed December 2004, URL `http://www.who.int/mediacentre/factsheets/fs282/en/`.

[75] A. J. Parker. Sense and the single neuron: Probing the physiology of perception. *Neurosci*, 21:227–277 (1998).

[76] S. Pastoor. Human factors of 3d imaging: Results of recent research at heinrich-hertz-institut berlin. Technical report, Heinrich-Hertz-Institut Berlin (1995).

[77] S. Pastoor. Human factors of 3d imaging: Results of recent research at heinrich-hertz-institut berlin. In *IDW'95*, volume 3, pages 69–72 (1995).

[78] R. Patterson and W. L. Martin. Human stereopsis. *Human Factors*, 34(6):669–692 (1992).

[79] M. Pauck. Experiences with the stereo world beamsplitter (1998). Accessed May 2003, URL `http://www.pauck.de/marco/photo/stereo/stereo_world/stereo_world.html`.

[80] Fatih M. Porikli and Richard V. Kollarits. Stereo image acquisition and display specifications for accurate depth perception. Technical report, Image Processing Department, ATT Research Labs (October 1997). URL `vision.poly.edu:8080/~fporikli/pdfs/depthperception.pdf`.

[81] C. Products. Image quality perception - evaluating real-world image quality performance, in clartity product marketing department. Technical report, Clartity (2002).

[82] N. Qian. Binocular disparity and the perception of depth. *Neuron*, 18:359–368 (1997).

[83] R.C. Thomas R. D. Boyle. *Computer Vision: A First Course.* Alfred Waller, first edition (1988). ISBN 0632015772.

[84] Martin Reddy. The graphics file formats page - 2d bitmap specifications (June 1997). Accessed Jan 2005, URL `http://www.dcs.ed.ac.uk/home/mxr/gfx/2d-hi.html`.

[85] T.O. Salmon. Introduction to binocular vision. Accessed March 2005, URL `http://arapaho.nsuok.edu/~salmonto/VSIII_2005/VSIII_2005.html`.

[86] D Scharstein and R. Szeliski. Middlebury stereo vision page: Data. Accessed June 2004, URL `http://cat.middlebury.edu/stereo/data.html`.

[87] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47:7–42 (2002).

[88] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:195–202 (June 2003).

[89] D. Scharstein and R. Szeliski. Middlebury stereo image page: Evaluation (2005). Accessed October 2005, URL `http://bj.middlebury.edu/~schar/stereo/newEval/php/results.php`.

[90] Daniel Scharstein and Richard Szeliski. Stereo data sets with ground truth. Accessed May 2003, URL `http://cat.middlebury.edu/stereo/newdata.html`.

[91] Daniel Scharstein and Richard Szeliski. Stereomatch. [Computer Program].

[92] Mel Siegel and S. Nagata. Just enough reality: comfortable 3-d viewing via microstereopsis. *IEEE Transactions on Circuits and Systems for Video Technology*, 10(3):387 – 396 (April 2000).

[93] Mel Siegel, Y. Tobinaga, and T. Akiya. Kinder gentler stereo. In *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems VI*, volume 3639A, pages 18 – 27 (January 1999).

[94] S.Pastoor and K.Schenke. Subjective assessments of the resolution of viewing directions in a multi-viewpoint 3d tv system. In *Proc. SID*, volume 30, pages 217–223 (1989).

[95] S.B. Andrei State and Henry Fuchs. Bunker view: Limited-range head-motion-parallax visualisation for complex data sets. In *Proc. Visualization in Biomedical Computing*, pages 301–306 (1994).

[96] Richard Szeliski. Prediction error as a quality metric for motion and stereo. In *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*, page 781. IEEE Computer Society (1999). ISBN 0-7695-0164-8.

[97] SourceForge Project Team. Gandalf. Accessed January 2003 [Computer Program], URL `http://gandalf-library.sourceforge.net/`.

[98] N.A. Valyus. *Stereoscopy.* Focal Press (1966).

[99] O. Veksler. Stereo correspondence by dynamic programming on a tree. *Computer Vision and Pattern Recognition* (June 2005). URL `dx.doi.org/10.1109/CVPR.2005.334`.

[100] H. Veron, D. A. Southard, J. R. Leger, and J. L. Conway. Stereoscopic displays of terrain database visualization. In *Proceedings of SPIE Stereoscopic Displays and Applications*, pages 124–135 (1990).

[101] T. Vilis. Physiology of the senses (undergraduate course) (2004). Accessed December 2004, URL `http://www.med.uwo.ca/physpharm/courses/sensesweb/`.

[102] All About Vision. Eye problems, injuries and diseases. Accessed June 2004, URL `http://www.allaboutvision.com/resources/statistics-eye-diseases.htm`.

[103] Dawn Vreven, Suzanne P. McKee, and Preeti Verghese. Contour completion through depth interferes with stereoacuity. *Vision Res*, 42:2153–2162 (April 2002).

[104] Nicholas Wade. *A natural history of vision.* MIT Press (1998). ISBN 0-262-23194-8.

[105] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4) (Apr 2004).

[106] Zhou Wang, Alan C. Bovik, and Ligang Lu. Why is image quality assessment so difficult? In *IEEE International Conference on Acoustics, Speech, & Signal Processing*, volume 4, pages 3313–3316 (May 2002).

[107] C. Ware, C. Gobrecht, and M. Paton. Dynamic adjustment of stereo display parameters. *IEEE Transactions on Systems, Man and Cybernetics—Part A: Systems and Humans*, 28(1):56–65 (Jan. 1998).

[108] Colin Ware. Dynamic stereo displays. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 310–316. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA (1995). ISBN 0-201-84705-1. doi:http://doi.acm.org/10.1145/223904.223944.

[109] Colin Ware. *Information Visualization: Perception for Design.* Morgan Kaufmann Publishers Inc. (Dec 1999). ISBN 1-55860-511-8.

[110] Zachary Wartell. *Stereoscopic Head-Tracked Displays: Analysis and Development of Display Algorithms.* Ph.D. thesis, Georgia institute of Technology (Aug 2001).

[111] Zachary Wartell, Larry F. Hodges, and William Ribarsky. Characterizing image fusion techniques in stereoscopic htds. In *GRIN'01: No description on Graphics interface 2001*, pages 223–232. Canadian Information Processing Society (2001). ISBN 0-9688808-0-0.

[112] Charles Wheatstone. Contributions to the physiology of vision.-part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transaction*, 128:371–394 (June 1836).

[113] J.S. Whiting, M.P. Eckstein, S. Einav, and N.L. Eigler. Perceptual evaluation of jpeg compression for medical image sequences. *OSA Annual Meeting Tech. Dig.*, 23(161) (1992).

[114] Wikipedia. Weight function. URL `http://en.wikipedia.org/wiki/Weight_function`.

[115] John Michael Williams. The pulfrich effect. Accessed Jan 2005, URL `http://www.siu.edu/~pulfrich/`.

[116] Steven P. Williams and Russell V. Parrish. New computational control techniques and increased understanding for stereo 3-d displays. *Stereoscopic Displays and Applications, Proceedings of the SPIE - The International Society for Optical Engineering*, pages 73–82. (1990).

[117] Graham J. Woodgate, David Erza, Jonathon Harrold, Nicolas S. Holliman, Graham R. Jones, and Richard R. Moseley. Observer tracking autostereoscopic 3d display systems. *Proceedings of SPIE, Stereoscopic Displays and Virtual Reality Systems IV*, 3015 (1997).

[118] Graham J. Woodgate, Jonathan Harrold, Adrian M. S. Jacobs, Richard R. Moseley, and David Ezra. Flat panel autostereoscopic displays - characterisations and enhancement. *SPIE*, 3957 (2000).

[119] Andrew Woods, Tom Docherty, and Rolf Kock. Image distortions in stereoscopic video systems. In *SPIE: Stereoscopic Displays and Applications IV*, volume 1915 (Feb 1993).

[120] Andrew Woods and John Merritt. Stereoscopic display application issues. *SPIE, Electronic Imaging 2004 Science and Technology Course* (Jan 2004). [Short Course].

[121] Yei-Yu Yeh and Louis D. Silverstein. Limits of fusion and depth judgment in stereoscopic color displays. *Hum. Factors*, 32(1):45–60 (1990). ISSN 0018-7208.