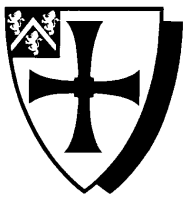
Stereo Display Systems

by

Andrew Thurman



Submitted in conformity with the requirements
for the degree of Electronic Engineering
Department of Engineering
University of Durham

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

Acknowledgements

Professor Purvis, my Supervisor and Tutor

The Mechanical Workshop, for producing the beam splitters

Richard McWilliam, for help with optical equipment

The volunteers, who took part in my experiments

The CImg C++ image library, by David Tschumperlé

Mike, Lesley and Mark Thurman, for proofreading

Jenny Radcliffe, for proofreading and LaTeX support

Edwin Brady, for the DUThesis LaTeX template

Abstract

Stereo Display Systems

Andrew Thurman

People have always wanted to obtain increasingly accurate sensory information. In the future it may be possible to record sensory inputs and replay them directly to the brain for a complete experience, but although that's not yet possible we can reproduce parts of the experience. Tactile sensory inputs are in their infancy, and olfactory inputs are not yet available. Surround sound can reproduce audio almost perfectly from life. This project focuses on the major human sense - sight - to raise the viewing of still images and video from a 2D image to a realistically 3D one.

Stereo imaging has been around for well over a century. Taking good quality stereo photographs, however, has only been possible for those with the correct equipment to take them. 'Home made' stereo images are difficult to align correctly and this misalignment will break down the stereo effect, often causing headaches.

This project shows that using Fourier transforms to align stereo images can improve usage for casual users, bringing headache free stereo images to the mass market. It also demonstrates that it is possible to greatly reduce file size with no noticeable loss of stereo effect, by reducing the amount of data repeated between images.

Original Gantt Chart

(Summer 07)	01/10/07	08/10/07	15/10/07	22/10/07	29/10/07	05/11/07	12/11/07	19/11/07	26/11/07	03/12/07	10/12/07
Research background to the project											
			The use of FFT to align stereo pairs								
				Design of a tripod extension							
						Image compression of still images - storing 1 + differences					
									Extending this to video images		
17/12/07	24/12/07	31/12/07	07/01/08	14/01/08	21/01/08	28/01/08	04/02/08	11/02/08	18/02/08	25/02/08	03/03/08
Christmas Vacation Continue to work on project											
				Buy 2 cameras of correct spec.							
					Run experiment on volunteers to test accuracy of stereo images						
							Write up				
10/03/08	17/03/08	24/03/08	31/03/08	07/04/08	14/04/08	21/04/08					
(Write up)											
Hand in first draft											
	Easter Vacation										
						Hand in final draft					

Contents

1 Introduction **1**

1.1 Alignment and Compression . . . 1

1.1.1 Alignment . . . 1

1.1.2 Compression . . . 2

1.2 Methods of depth perception . . . 2

1.2.1 Binocular disparity and convergence . . . 2

1.2.2 Accommodation . . . 3

1.2.3 Motion parallax . . . 4

1.2.4 Occlusion . . . 6

1.2.5 Lighting . . . 7

1.2.6 Atmospheric . . . 7

1.3 Binocular disparity on a virtual screen . . . 8

2 Previous Work **11**

2.1 Producing . . . 11

2.1.1 Taking two images using a beam splitter . . . 11

2.1.2 Taking two images with synchronised cameras . . . 12

2.1.3 Taking two images asynchronously with the same camera . . . 12

2.1.4 Alignment . . . 13

2.2 Storing and Transmitting . . . 13

2.2.1 Stereo formats and compression . . . 14

2.2.2 Standard image compression types . . . 15

2.2.2.1 Lossy compression . . . 15

2.2.2.2 Lossless compression . . . 16

2.3 Displaying . . . 16

2.3.1 Autostereoscopic (Free viewing) . . . 16

2.3.1.1 Stereograms . . . 16

2.3.1.2 Ivanov projection . . . 17

2.3.1.3 Lenticular or Fresnel lens . . . 18

2.3.2 Stereoscopic (Non-free Viewing) . . . 18

2.3.2.1 Anaglyph / Colourcode . . . 19
2.3.2.2 LCD shutter glasses . . . 19
2.3.2.3 Polarised glasses . . . 20
2.3.2.4 Pulfrich glasses . . . 20
2.3.2.5 Chromadepth glasses . . . 21

3 Theory **22**
3.1 Alignment . . . 22
3.2 Compression . . . 26
3.2.1 Standard image compression . . . 26
3.2.2 Only store partial FFT . . . 26
3.2.3 Phase correlation peaks . . . 26
3.2.4 Manual pixel comparisons . . . 27
3.2.5 Potential overlays . . . 27
3.2.6 Final solution . . . 28
3.3 Beam splitter . . . 29
3.3.1 Synchronised camera stand . . . 33

4 Experimentation **35**
4.1 Experimental results . . . 37
4.1.1 Repeated images . . . 39
4.1.2 Changes over time . . . 39
4.1.3 Compressed file sizes . . . 40
4.1.4 Compression . . . 41
4.1.5 Alignment . . . 42
4.2 Beam splitters . . . 43

5 Discussion **45**
5.1 Alignment . . . 45
5.2 Compression . . . 46
5.3 Beam splitter . . . 49

6 Conclusion **50**
6.1 Alignment . . . 50
6.2 Compression . . . 50
6.3 Beam splitter . . . 50

List of Figures

1.1 Binocular Disparity . . . 3
1.2 Accommodation . . . 4
1.3 Motion Parallax . . . 5
1.4 Occlusion . . . 6
1.5 Atmospheric fading . . . 8
1.6 Binocular disparity and convergence on a virtual screen . . . 9
1.7 Depth field of the DTI Virtual Screen . . . 10

2.1 Ivanov Projection . . . 17
2.2 Lenticular lens . . . 18
2.3 Comparison between Lenticular and Fresnel lenses . . . 19
2.4 Polarisation . . . 20

3.1 Offset axes . . . 22
3.2 Argand diagram showing phase correlation . . . 24
3.3 Final compression method . . . 30
3.4 Beam splitter (first design), top view . . . 30
3.5 Beam splitter (first design), intended rear view . . . 31
3.6 Beam splitter (first design), actual rear view . . . 31
3.7 'Shoebox' or Tunnel effect inherent in the beam splitter . . . 32
3.8 Beam splitter (second design), top view . . . 33

4.1 Graph of experimental quality judgements over time . . . 39
4.2 Graph of file size against compression . . . 40
4.3 Graph of image quality against compression . . . 41
4.4 Graph of image quality against alignment . . . 42
4.5 Layout for the beam splitter experiment . . . 43
4.6 Results from the beam splitter experiment (1) . . . 43
4.7 Results from the beam splitter experiment (2) . . . 44

5.1 Stereo pair, pre-compression, with non-horizontally repeated data . . . 47
5.2 Stereo pair, post-compression, with non-horizontally repeated data . . . 47

List of Tables

4.1 Table of compressed file sizes 37
4.2 Summary table of results for all image types 38
4.3 Table of variation between repeated images 39

Glossary of Terms

Nomenclature and Acronyms

2D Two dimensional

3D Three dimensional

G_A The result of the 2D Fourier transform on image A

GIF Graphical Interchange Format

h Image height

JPEG Joint Photographic Experts Group; also the file format developed by them

JPS JPEG Stereo format

LZW 'Lempel - Ziv - Welch' - a compression algorithm the name of which is derived from the initials of its creators

MLA Machine Learning Algorithm

mm Millimetres

MPEG Motion Pictures Experts Group; also the file formats developed by them

PNG Portable Network Graphics

R Rotational offset

T Threshold value

VESA Video Electronics Standards Association

w Image width

X Horizontal offset

Y Vertical offset

Z Offset perpendicular to the screen (zoom)

Miscellaneous Terms

cardboarding Horizontal image offset is less than the inter-ocular distance; objects will appear wider, higher and shallower than in reality

depth cue A means by which the brain can detect depth in a three dimensional scene

DTI2015XLS The Dimension Technologies Inc. Virtual Screen, a stereo display screen used in this project for testing

ghosting or crosstalk Where each eye is able to see both halves of the stereo pair, as opposed to each seeing just the correct image alone

hyperstereoscopy Horizontal image offset is greater than the inter-ocular distance; objects will appear smaller but deeper than in reality

keystoning or toe-in Where an image is taken not perpendicular to the plane being photographed, leading to minor distortion

offset The movement required in an axis (X, Y, Z, R) to change the first image of a stereo pair to be near identical to the second

side:side Stereo pair are laid out beside each other

top:bottom Stereo pair are laid out on top of each other

Chapter 1

Introduction

The Ancient Greeks were the first to document depth perception, when Euclid wrote his seminal work on The Optics, and over the two thousand years since its discovery humans have been becoming ever closer to reproducing the three dimensional world around them. Leonardo da Vinci - following on from Euclid's work[1] - was one of the first to create paintings with accurate perspectives, but it was not until Charles Wheatstone's Reflecting Stereoscope[2] in 1838 that people were able to see genuine stereo images.

1.1 Alignment and Compression

The main focus for this project has been on the alignment of stereo images and their compression.

1.1.1 Alignment

Stereo imaging has been around for well over a century, and has taken a variety of forms. Stereo viewers, showing professional stereo images, have been popular since the middle of the 19th Century, and became a common entertainment in homes. Taking good quality stereo photographs, however, has only been available to those with the correct equipment to take them. Methods have included stereo cameras and devices for beam splitting, but 'home made' stereo images (i.e. using a standard camera to take the two images) are difficult to align correctly and this misalignment will break down the stereo effect, potentially causing headaches.

If a method could be devised to realign two images taken asynchronously by a single, unmounted camera then it would be possible for anyone to start to take stereoscopic photographs, opening stereo photography to the casual home user.

[1] Wade, Ono and Lillakas 2001
[2] Wheatstone 1838; Silverman 1993

1.1.2 Compression

The main focus of experimentation in the field of stereoscopic imagery has been concentrated upon the production of better images, producing many methods both for recording images and for displaying them. The one element that has had little changed is the storing and transmission of the images. Stereo images are typically stored as two separate pictures - one for each eye - which makes a stereo image twice the size of a standard monoscopic, two dimensional one. This is only made worse by the advent of stereo video, with far larger file sizes that are then doubled. This is highly inefficient, as both pictures making up the stereo image are looking at the same objects in the same scene. This means that the images will contain a lot of data that is identical.

If a method could be devised to only store the data that had changed then it would greatly reduce the size of stereo files, and thus greatly increase their transmission speed.

1.2 Methods of depth perception

Humans perceive the world around them in three dimensions. This is predominantly due to the stereo effect which can be created with two forward facing eyes, but this is by no means the only method used by the brain to calculate distance, depth and layering.

1.2.1 Binocular disparity and convergence

Binocular disparity is in many ways the 'basis of stereopsis'[3] - the main method by which the brain can detect the third dimension. The images perceived by each eye contain slight differences, where objects are offset differently relative to the image dependent on their depth in the scene. The two images of each object will then only fuse into one, three dimensional image when centred to both eyes - i.e. only when both eyes are aimed directly at that object's depth (see Fig. 1.1). The brain can then detect depth by the convergence (amount of angling inwards) of the eyes, where objects requiring great convergence are close and objects requiring little or no convergence are distant. This is therefore mirrored in the images perceived by the eyes, allowing the brain to calculate relative depths without having to necessarily focus on each object in the image. That is, if an object appears in the centre of both images, then its depth is at the depth of the convergence. If it appears to the right of the left image and the left of the right image (negative disparity - towards the centre, when placed side:side), then its depth is less (it is closer) than the depth of the convergence. If it appears to the left of the left image and the right of the right image (positive disparity - towards the outside, when placed side:side) then its depth is greater (it is further away) than the depth of the convergence.

Note that in real life it is not possible for the two images of an object to not have fused at any convergence, as the eyes can see to infinity when parallel and thus no object should give

[3] Purves, Augustine, Fitzpatrick, Lawrence, Lamantia and McNamara 1996

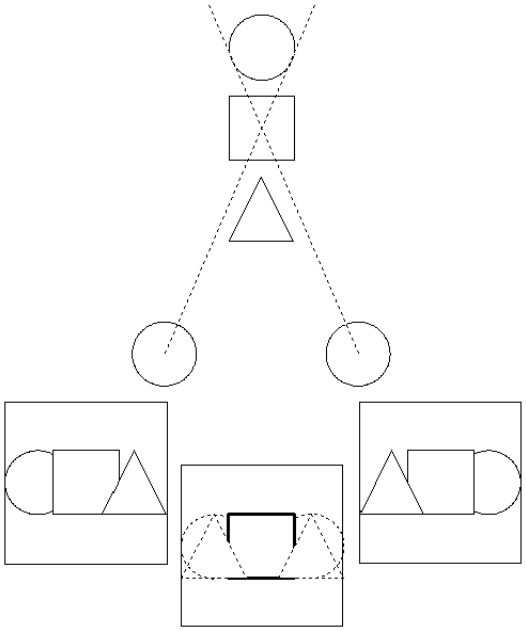


Figure 1.1: Binocular Disparity - the eyes will fuse on the object centred in their vision, doubling objects at different depths.

a greater disparity than that - i.e. the eyes should never have to become divergent (pointing away from each other). When creating stereo images it is possible to create the effect where an object has a greater disparity than that of an object at infinity. As the human brain is not equipped for this it causes a severe break down in the stereo effect and can lead to headaches.

The eyes are only able to perceive binocular disparity in a limited field - a three dimensional space in front of the eyes inside of which objects will fuse to form a three dimensional image and outside of which objects will appear doubled (as two images which the brain cannot fuse). The area is known as Panum's Fusion[4], and its bounds are set by the convergence of the eyes such that the objects on which the eyes are centred will mark the centre of Panum's Fusion. As it is governed by convergence it poses little difficulty to most forms of stereo display system, which merely give the offsets to the eyes and allow for convergence as part of binocular disparity.

1.2.2 Accommodation

While binocular disparity and convergence are highly important in depth perception, each eye is also capable of calculating depth independently via accommodation. Accommodation is where each eye will alter the curvature of its lens to bring the image seen by it into focus. A greater curvature means that the eye is focussing on an object that is very close, while a

[4]Diner and Fender 1988

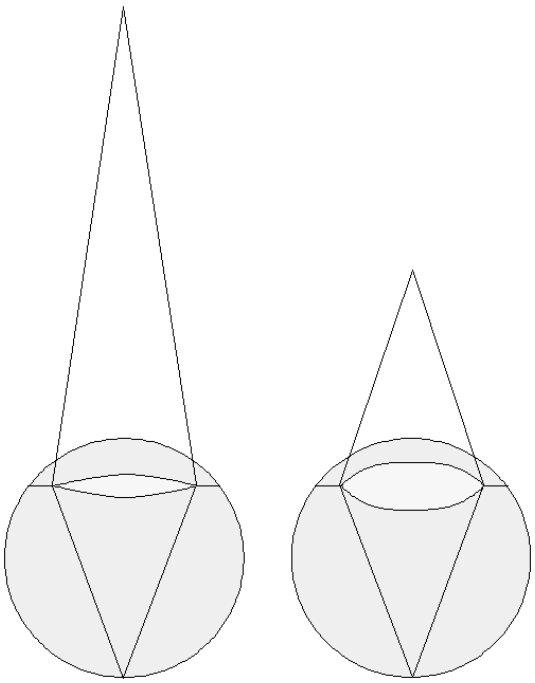


Figure 1.2: Accommodation - the eye will change the curvature of its lens in order to alter the focus.

lesser curvature means that the eye is focussing on an object that is at a distance. While this is not as accurate as binocular disparity the brain is able to use accommodation to calculate rough depths as well as relative depths from how out of focus one object is compared to another - i.e. while looking at a close object, the brain expects distant objects to become more blurred the further they are from the object of interest, and similarly in reverse when looking at a distant object.

Accommodation is one of the few depth cues available to the brain that cannot be fully tricked by artificial means[5], although it may be possible in the future to create a system which would modify the displayed images in response to changes in lens curvature[6]. In practice, users of stereo display systems of various types - whether head mounted displays, stereoscopic screens or VR shutterglass environments - quickly learn to ignore the depth cues given by accommodation, as they are contradictory to the other cues given to them by the system used.

1.2.3 Motion parallax

The further away from the viewer an object is, the less the disparity between the perceived position of its image at different viewpoints will be relative to the distance between

[5] Cruz-Neira, Sandin and DeFanti 1993
[6] Wann, Rushton and Mon-Williams 1995

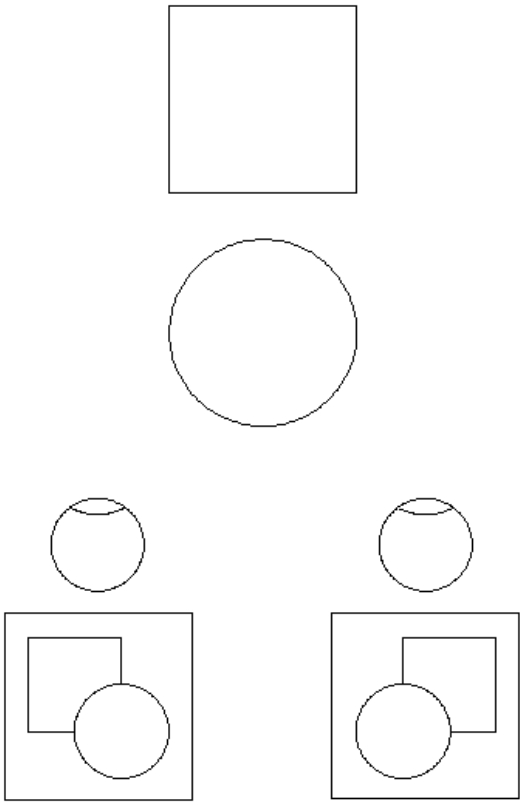


Figure 1.3: Motion Parallax - close objects will move further across the field of view during a horizontal displacement than distant objects.

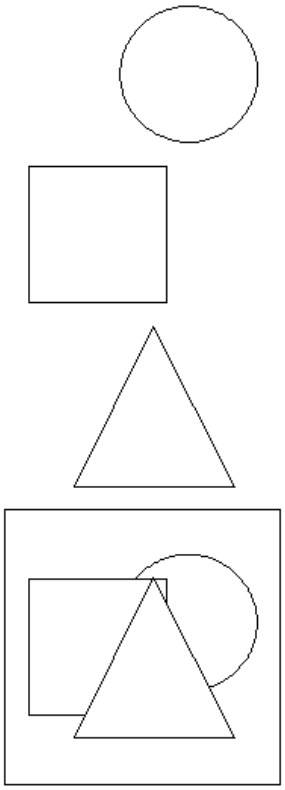


Figure 1.4: Occlusion - Close objects that are in the same line of sight as distant objects will partially hide them from view.

viewpoints. In other words, a close object will move a large amount for a small change in viewpoint, while a distant object will barely move at all. This is the basis for binocular disparity, where the two viewpoints are the eyes and the distance between them the inter-ocular distance, and the brain also uses it on a larger scale to calculate distances and depths more accurately by moving the eyes horizontally. As it uses the same basic principle as binocular disparity, which is as stated above the primary means of depth perception in humans, it is very accurate and the brain is used to using it for calculating depth.

Motion parallax does, however, require the ability to change the point of view of the person, which reduces its possible use in displays which do not take into account the position of the viewer relative to the display.

1.2.4 Occlusion

Binocular disparity, convergence, accommodation and motion parallax can not usually be used when viewing still images on standard, two dimensional displays with no head tracking (where the display detects the position of the user relative to the screen). There are several visual depth cues, however, which do not require multiple viewpoints (either head

moving or stereo imagery) or actual depth. These give depth to flat displays, paintings and photographs. One of these depth cues is occlusion.

Occlusion is a depth cue which does not require stereo vision and is a major reason why two dimensional pictures can appear to have depth. If part of an object is hidden because part of another object is in the way, it can be assumed that the object covering is closer to the viewer than the object being covered. While this does not allow for relative depths (it is impossible, through use of occlusion alone, to tell whether one object is a centimetre behind the other or a mile) it does let the brain calculate the layering of objects such that in a busy scene (one with many overlapping objects) it can tell which objects are in front of which.

Occlusion is the method of creating depth which is easiest to recreate, under any system. It is highly effective (to the point that lack of occlusion would render a system unusable), can be produced on a flat, two dimensional display, and has been used in pictures for millennia.

1.2.5 Lighting

Another means by which the human brain can work out the relative distances of objects is by plotting their shadows. In most lights objects will cast a visible shadow (this includes surface shading, where the side facing away from a light source will be darker than the side facing towards it). The position of the shadows could be easily calculated from the known positions of the object and the light source, and in reverse therefore the object position can be easily calculated from the known position of the shadows (through occlusion - which objects block the shadow, and which are shaded by it) and of the light source (which is similar for all objects, and can be calculated from the shadow positions).

Similarly to occlusion, recreating this depth cue does not require complex hardware and has been used for a long time in paintings to add realism and depth. It is such a descriptive yet simple method that most three dimensional renders will contain raytracing to give realistic light and shade.

1.2.6 Atmospheric

There are few depth cues that have not already been covered. One that is only occasionally applicable is that of atmospheric conditions, such as fog. If the visibility of objects is proportional to their distance from the viewer (such as in a fog or mist where objects fade from view over distance) then the brain can use this to calculate relative distances. This is based on current depth knowledge - i.e. if the brain knows the depth of a certain object and knows how much it has faded over that distance then it can calculate the distance of an object which has faded over a different distance.

This can be a difficult one to recreate in the context a three dimensional display, purely because it is an effect that is not often in evidence. It is only useful when creating a scene which would contain atmospheric effects - such as outdoors in the fog or mist.

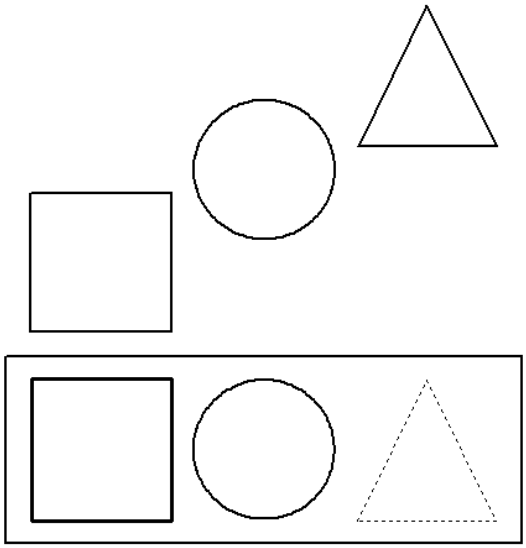


Figure 1.5: Atmospheric fading - In certain atmospheric conditions (e.g. fog) objects will appear to fade from view the further they are from the viewer.

1.3 Binocular disparity on a virtual screen

Most stereo display systems (see 2.3) make use of binocular disparity (see 1.2.1) to allow the viewer to see objects that are behind or in front of a screen. This does have limitations, however, as to the depth of the screen and the amount that objects can come forwards.

The screen will be showing two instances of every object, one for each eye (see Fig. 1.6). If the object appears to lie in the plane of the screen then both will overlap and the eyes will focus on the screen (Fig. 1.6 B). If it is appearing to lie behind the plane of the screen then the two images will be diverged and the eyes, focusing on the images given to them, will converge at a point behind the screen (Fig. 1.6 C). If the object is appearing to lie in front of the screen then the two images will have crossed over so that the eyes will converge at a point in front of the screen (Fig. 1.6 A).

In order for the two images making up an object's stereo pair to fuse into a coherent image they must both be fully visible. While the brain can cope with seeing only one instance of an object (for example when looking round a corner, where one eye can see an object but the other has a reduced field of view due to the corner), it can lead to a partial breakdown in the stereo effect if the object is partly hidden 'behind' the screen (Fig. 1.6 E). If the object is in front of the screen then the breakdown is total, as the brain cannot find a reason for one eye being blocked (Fig. 1.6 F).

These constraints allow for a three dimensional field of view to be drawn (see Fig. 1.7). As the inter-ocular distance is known (approx. 70mm for the average Western adult[7]),

[7] Jeong and Bjelkhagen 1992

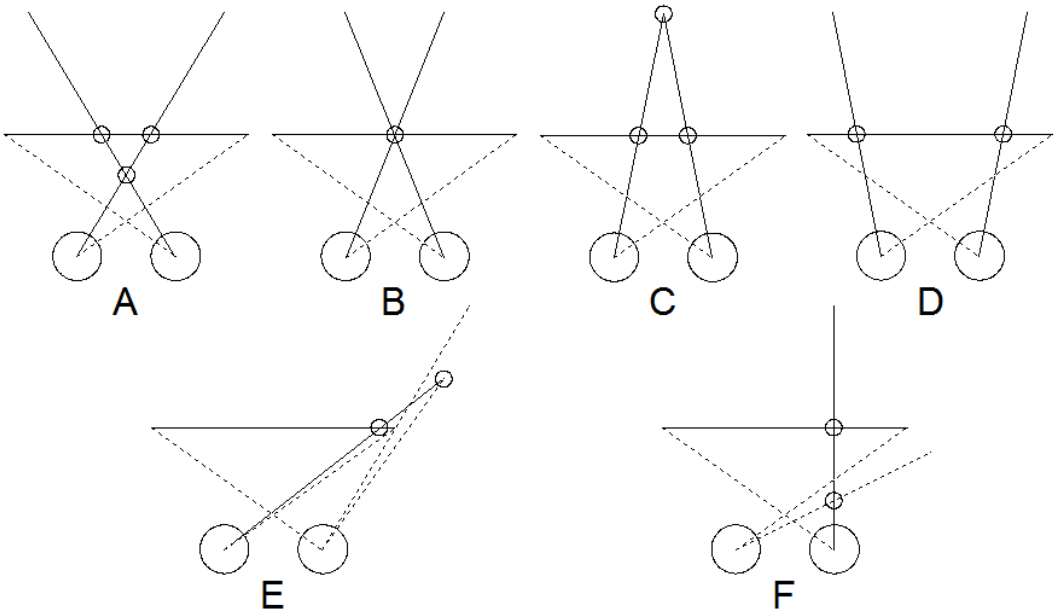


Figure 1.6: Binocular disparity and convergence on a virtual screen:
A - Convergent images; the eyes will focus on a virtual object in front of the screen.
B - Images overlap; the eyes will focus on a virtual object in the plane of the screen.
C - Divergent images; the eyes will focus on a virtual object lying behind the plane of the screen.
D - Over divergence, beyond parallel; the eyes will fail to converge at any point, and the brain will fail to fuse the image.
E - One half of the stereo pair for an object behind the screen is hidden off screen; the brain can cope with this, as there is an object (the edge of the screen) behind which it is hidden. There is a partial breakdown in the stereo effect.
F - One half of the stereo pair for an object in front of the screen is hidden off screen; the brain will be unable to process this as there is no object for it to be hidden behind and so a breakdown in the stereo effect is inevitable

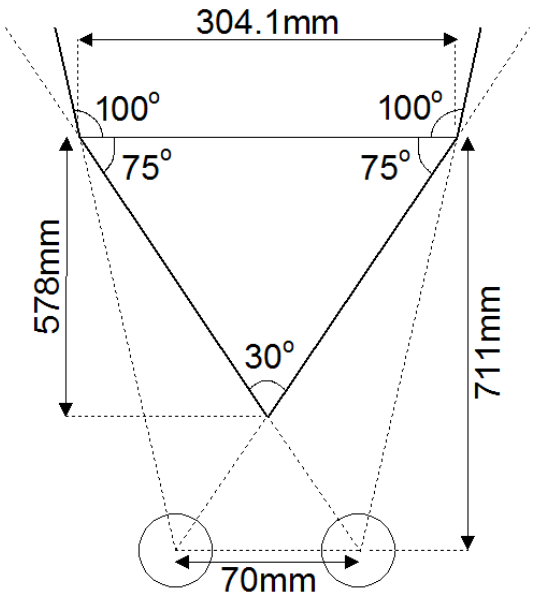


Figure 1.7: Depth field for the DTI 2015XLS Virtual Window from Dimension Technologies Inc. Objects will only be fully visible when fully enclosed within the field.

and the screen dimensions (304.1mm x 228.1mm[8]) and optimum screen distance (711mm ± 102mm[9]) are known, then simple geometry can be used to calculate the size of the viewable three dimensional area.

Note that in some cases stereo displays may in practice have limited depth ranges, beyond which objects may appear distorted. This can be as little as 55mm from the screen plane[10].

[8] *User Manual for the DTI 2015XLS Virtual Window*^{TM} 2001
[9] *User Manual for the DTI 2015XLS Virtual Window*^{TM} 2001
[10] Holliman 2004

Chapter 2

Previous Work

The field of stereo imaging can be split into three discrete sections:

- Producing the stereo pair
- Storing and transmitting the stereo pair
- Displaying the stereo pair

These are all, to a great extent, separate - the decision as to which method to use for producing is independent of that for the storing and transmission and that for the displaying.

2.1 Producing

There are many methods that have been used to take stereo images in the past, but these broadly fall into three main categories, each with associated problems.

2.1.1 Taking two images using a beam splitter

To take two images onto the same 'picture' with one camera (aligned either side:side or top:bottom) it is necessary to have an arrangement of lenses or mirrors known as a beam splitter. This must be either aligned with the camera or mounted onto it, and may be bulky, expensive to produce, and fragile. The images each take up half of the resolution of the camera, reducing the image quality of the stereo pair. It can, however, produce the best images from all three categories, in terms of accuracy of alignment between the images. The images may be offset from the picture in which they are contained (if there is misalignment between the camera and the beam splitter), but they will always be perfectly aligned with each other in Y, Z and R while at the optimum X offset.

2.1.2 Taking two images with synchronised cameras

The main problem with taking two images with two synchronised cameras is that it forces the user to carry two cameras. While this may not be a large difficulty - when, for example, compared to carrying a beam splitter - it does require the user to own two cameras, and for the two cameras to be of as near identical specification as can be reached. The image settings (zoom, light balance, etc.) must also be identical on both cameras for each pair of photographs taken.

It is preferable, therefore, to have the two cameras attached to a double mount which can hold them at the correct horizontal offset and fix their angles and vertical offsets to be equal. The double mount is not essential, but if not used then the only advantage of this method over the final category (one camera taking both images asynchronously) is the synchronicity - i.e. moving objects will be in the same relative position between the two images.

2.1.3 Taking two images asynchronously with the same camera

The two images can also be taken asynchronously using one camera, which is used to take the first image and then manually moved to take the second. The great advantage of this is that it is possible for any person with a camera to take stereo photos, without the need for either specialist equipment or a second camera - thus enabling most people to start photographing stereo images instantly and greatly reducing the cost of equipment for people who do not already possess a camera.

The disadvantages, however, are that there is no guarantee that the two images are aligned in any way - in any of the X, Y, Z or R axes (see 3.1), in convergence (i.e. images may be parallel, toed-in (converging) or divergent - see 1.2.1), or (importantly, if taking images other than still life or scenery) in time. The alignment problems are potentially solvable, however. Changes in offset in X, Y, Z and R can be calculated; problematic convergence can be reduced if the camera is aimed at the same specific object towards the horizon in both images (introducing slight toe in if the object is close, but reducing overall convergence errors); and changes in time can be ignored if the temporal differences are negligible (for example in the case of two images being taken in quick succession, or a scene with little or no foreground motion).

Note that the optimum separation of the two images, in all three cases, is 70mm - this being the average inter-ocular distance for a human. If the horizontal image offset is greater or smaller than this then the person viewing the images will be given the effect of having his eyes either further apart or closer together than they actually are. The main effect caused by this is to change the perception of the scale of the scene - both in terms of horizontal and vertical scale and inversely of depth. If further apart than 70mm the effect is known as hyper-stereoscopy, where all objects in the scene appear to be far smaller than they actually are but objects will look much deeper. If closer together the effect is known as ‘cardboarding’,

where although objects appear to be larger than they actually are foreground objects are made to appear thinner, to the extent that in the extreme objects can appear to be two dimensional - or cut out of cardboard (giving the name 'cardboarding').

For the most part, therefore, it is necessary to keep the horizontal image offset at the inter-ocular distance, although there are a few specialised cases where this is not the case (for example, when photographing a distant and very large object hyper-stereoscopy can be recommended - it reduces the size of the object, but increases the depth of distant objects that would normally appear flat in comparison to their distance from the viewer)

2.1.4 Alignment

Once the stereo images have been produced they may be misaligned (especially if taken using the third method of taking both images asynchronously with one camera), and this project was in part to investigate a method of realignment.

A process of alignment between two images (Image Registration[1]) also has many other uses, ranging from the field of medical imaging to aerial photography. It is normally used to allow several images of the same object, taken at different times or from different angles, to be aligned for ease of viewing. Photographs of overlapping areas of ground can be aligned to produce a single image of the whole area, which would be impossible to take in one image. The same process can be used when scanning large documents in in parts, or to produce a panoramic shot from a series of smaller photographs. Image registration is also of great importance for applications such as fingerprint recognition, where the test print needs to be aligned with and tested against the image held on file.

Alignment is achieved by calculating the offsets between the two images and translating one image by the offset relative to the other. The offsets can be calculated in either the spatial or frequency domain, although in almost all cases the frequency domain is chosen. This gives a far more efficient set of algorithms[2] which can make use of Fourier transforms, allowing for Fast Fourier Transforms to further increase the speed.

2.2 Storing and Transmitting

Typically, most stereo systems have stored and transmitted the two images separately - either as two physically separate channels, or separately on the same image. This can take several different forms. Side:side and top:bottom are where the two images are simply placed next to each other, either left and right or top and bottom. Interlaced is where alternate lines are taken from each image, either horizontally or vertically. Frame sequential is more applicable to video than to still images, as it is where alternate frames are taken from each stream.

[1] Brown 1992
[2] Haering and da Vitoria Lobo 2001

2.2.1 Stereo formats and compression

There have been previous attempts at creating a format for stereo images and stereo video; either as a simple container format (where two files are saved as one with no compression) or with attempted compression[3].

One important quality to take into account when compressing or storing stereo images is that of backwards compatibility. In the majority of cases, a stereo image can be seen as being a two dimensional image with added depth. As such it would be useful to have the capability of viewing the stereo image as a flat two dimensional image on a two dimensional display - in cases where the user does not have a stereo display available, but still wishes to see the content of the image. It is analogous to viewing a colour image on a black and white display - a large quantity of data is lost (in this case the colour; in the case of stereo formats the depth), but enough remains for it to be usefully viewable to the user. One common format is JPEG Stereo (extension .jps), where the two pictures that make up the stereo image are saved as two halves of a flat (i.e. two dimensional) .jpeg file. A bit is then set to specify the orientation of the two images (whether side:side or top:bottom, and which is which), and the extension changed. A .jps reader will then be able to present the two halves of the image in the correct way for the viewing device used. This format works well in that it allows for images to be stored in a common and readily available (although lossy) format, in itself does not modify the images, and can be read by a standard .jpeg viewer if a .jps viewer is unavailable. The two disadvantages of the format are that the use of .jpeg compression means that the format is not lossless (see 2.2.2) (although an identical but lossless format specification could be created using .png), and that the only compression used is the jpeg compression on the individual images - there is no additional compression to make use of the similarities between the images.

The same basic method (side:side or top:bottom) can also be used for storing stereo video, while a further method is to interleave alternate frames of left and right images. This shares many of the advantages of the .jps format in that it stores the stereo video in effectively the same format that would be used for a two dimensional video and does not modify the frames, but there are still disadvantages - again there is no compression, keeping file sizes large, and although it can still be viewed on a standard display the resulting video would have a doubled effect as the brain perceived both images simultaneously through both eyes. Nevertheless, the Video Electronics Standards Association (VESA) does specify a standard for a synchronisation signal to allow the viewer to differentiate between left and right eyes.

Neither of these methods include compression to reduce the repeated data between the two images. There have been attempts to provide a means to do so, such as the compression scheme proposed by Mark Rosamond[4]. This was based on the principle that that if the

[3]Roberts and Slattery 2000
[4]Rosamond 2004

difference between the same pixel in the two images is below a set threshold then the brain will not notice if they are shown as the same pixel. This reduces the number of pixels that need to be transmitted, but while it compares the two images against each other it does not take into account the actual offsets between the two images. The pixels being matched are unlikely to be representations of the same point in three dimensional space, unless they are part of the background (far enough away from the camera that the stereo effect is negated). The method will still be effective, however, if it is assumed that objects will be relatively uniform (i.e. while the same point in three dimensional space is not being compared the pixels may be sufficiently similar, especially if part of the same object, for the difference to be neglected). As it did not take into consideration the actual offsets, however, the compression accuracy was highly dependent on the content of the image. Other methods making use of the repeated data include that put forward by Julien Flack et al[5], which utilises machine learning algorithms (MLAs) to allow a computer to calculate the changes in offset between the two images. This is, however, liable to be memory intensive - and as such there is plenty of scope for a simpler method for compression.

2.2.2 Standard image compression types

Far more work has been undertaken to provide methods for compressing standard, two dimensional images than has been to compress stereo images. These fall into two categories - Lossless and Lossy.

2.2.2.1 Lossy compression

Compressing data means fitting the data into less space. One way of doing this is to reduce the amount of information held by that data. In other words, the image recreated when decompressed will not be identical to the original image before compression. The decompressed image will be of lower quality, as information has been removed, but (assuming a suitable algorithm has been used for compression) the missing information is the information that is least noticeable.

The reason for using lossy compression is that in a large number of cases a lossy compression can give a far greater reduction in size than a lossless one, while retaining sufficient information that the brain cannot tell that the information is missing. The brain is very adept at filling in missing information, to the extent that an image can have changed by a large amount - in terms of actual pixel values, shadings, colours, etc. - while to the person looking at it the changes seem negligible.

The higher the compression the lower the quality of the image will fall, however. Defects in the image become more and more noticeable to the human eye, and are known as compression artefacts. Compressing a file multiple times will reduce the quality for each iteration, as more information will be lost through each compression.

[5]Flack, Harman and Fox 2003

Typical examples of lossy compression include the JPEG algorithm and the various MPEG algorithms.

2.2.2.2 Lossless compression

Lossless compression, as its name suggests, is used to compress data without losing any information, such that the decompressed image will be identical to the original image. Lossless compression works by storing the same information in less data, rather than reducing the amount of information. More probable segments of data (the probability being calculated from previous data, which may change through the course of the compression) are encoded in shorter 'codes' than less probable ones, so that the overall length is reduced.

While lossless compression gives exact images that lose no information in compression, the downside of using them is that they will not compress to as small a file as is probable from lossy compression (although this is partly dependent on image content).

Typical examples of lossless compression include the PNG format (based on the Deflate algorithm) and the GIF format (based on the LZW algorithm).

2.3 Displaying

There are several devices and methods for displaying stereo pairs. Here are outlined some of the more common methods.

Note that this does not include methods for displaying other types of three dimensional images, such as holographic, swept surface or static volume displays, or pseudo three dimensional images, such as projection onto a curved screen, as these do not display stereo images (see Appendix, section 1).

Stereo image displays fall into two categories - autostereoscopic (or free viewing), and stereoscopic (or non-free viewing).

2.3.1 Autostereoscopic (Free viewing)

These devices do not require the viewer to wear or use any special hardware to see the stereo image, but typically have the disadvantage that very few people can view them at any one time as their eyes must be in the correct locations to see the two images making up the stereo pair.

2.3.1.1 Stereograms

These are two side:side displayed images making up a stereo pair. The viewer then changes the convergence of his eyes until they fuse to form a three dimensional image - either by making his line of vision parallel (such that the left image is centred in the left eye and the

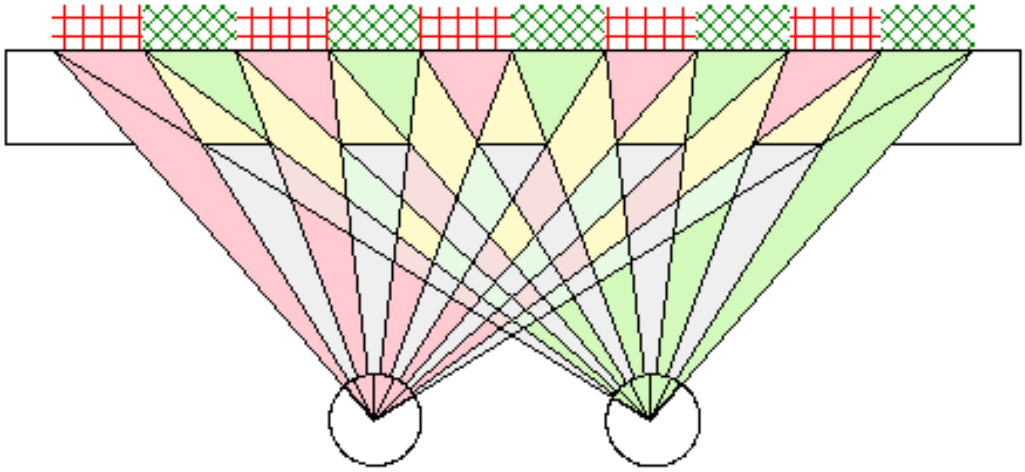


Figure 2.1: Ivanov Projection. The parallax barrier in front of the main display only allows alternate vertical lines to be visible to each eye (when the eyes are positioned in the correct places)

right image is centred in the right eye), or by crossing his eyes (such that the left image is centred in the right eye and the right image is centred in the left eye).

The difficulty with this method is that the viewer must be trained to use it - changing the eye convergence does not come naturally. It does, however, allow for several people to view the images at once - making it one of the few autostereoscopic methods to do so.

2.3.1.2 Ivanov projection

The Ivanov projection was first designed by D.V. Surenskii and S.P. Ivanov for Soyuzdetfilm, in Russia in the 1940s[6]. It makes use of a secondary screen or parallax barrier at a set offset in front of the display screen, which is purely a series of vertical lines with a fixed width and separation (see Fig. 2.1). The stereo image is then shown interlaced on the display screen, with alternate lines being taken from either the left or the right image of the stereo pair. The vertical lines on the parallax barrier in front will, when seen from a specific angle and viewpoint, align with the even lines on the rear display. As these hidden lines will be dark and the visible lines will be bright, it will appear to the brain to be a half resolution image composed solely of the alternate visible lines. The same vertical lines on the parallax barrier will, at a different angle and viewpoint, align with the odd lines - this time showing the viewer the other image. If the viewer sits in the correct position then each eye will be able to see a whole and different image, which make up the stereo pair.

This same method is being used with LCD screens to produce free viewing stereo displays for computers and televisions. One example of this is the DTI 2015XLS Virtual Window,

[6]Lentjes 2006

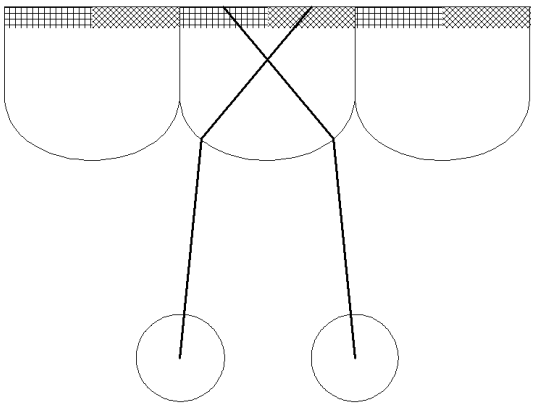


Figure 2.2: The lens bends the light from the two images, allowing the alternate lines to be viewed by each eye

from Dimension Technologies Inc. (although this does also make use of lenticular lenses), which was used over the course of this project to test and view stereo images. The main downfall of this method is that the optimal viewing position can be minimal.

2.3.1.3 Lenticular or Fresnel lens

This gives a similar effect to the Ivanov projection, but through a different means. Instead of having a secondary raster screen, a ribbed sheet of half cylinders (Lenticular) or hemispheres (Fresnel) is placed in front of the interlaced images. These focus the light in such a way that even lines are shown angled to one side and odd lines are shown angled to the other. Again, the viewer must sit in the correct place to see the stereo effect.

The practical difference between using lenticular and Fresnel lenses is that the Fresnel lens will give a greater viewing angle (a wider area in which the stereo effect can be seen), but with a reduction in image resolution. This is because the lenses (of both types) cannot channel their full radius width at the correct angles. There are therefore sections of the screen behind the lenses which will not be visible due to the position of the lens in front of it. With the Fresnel lenses there will be more lenses for a given resolution of display, and the hemispheres do not tessalate as the half cylinders do, further wasting the data. This reduces the effective resolution transmitted to the eyes.

2.3.2 Stereoscopic (Non-free Viewing)

While autostereoscopic systems can be good from both a user's point of view and from the perspective of the person producing the display (as there is no necessity to produce potentially expensive glasses for possibly a large number of people, or for the people to wear them), there is the difficulty that they cannot be easily seen by many people simultaneously.

Stereoscopic systems bypass the problem of correctly focusing the images directly to the

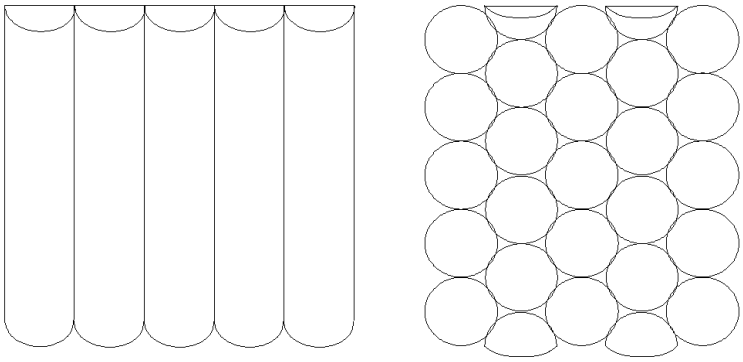


Figure 2.3: Lenticular and Fresnel lenses. The lenticular lens will have a higher resolution, while the Fresnel lens will give a greater viewing angle.

eyes by creating a filter that will only allow the correct image through for each eye. This makes them good for mass audiences, but often at a higher cost.

2.3.2.1 Anaglyph / Colourcode

The simplest method of filtering the two images is to restrict the frequencies of light able to pass through. Anaglyph and Colourcode both do this by placing coloured filters over the eyes, allowing through monochromatic images of the correct colours. Anaglyph uses Red/Blue, while Colourcode uses Yellow/Blue. While this effectively separates the images, it does lose the colour information. This can be re applied to anaglyph images by adding green, although this can potentially cause ghosting (the wrong image becoming visible on the other view) if care is not taken. Colourcode, being (Red+Green)/Blue, is an attempt to do the same.

One potential downfall of this method is that to create the coloured monochromatic images it is impossible to reach white as white would contain both colours and thus pass through both filters. The maximum brightness is therefore 50% in the colour - i.e. true red or true blue, for anaglyph In practice the brain will ignore this - as there are no brightnesses greater than 50% it will automatically compensate to make that appear as white. A far greater problem is ensuring that the display has exactly the same shade of colour as the glasses, otherwise ghosting and crosstalk can result.

2.3.2.2 LCD shutter glasses

Another option is to change the display between images at a set refresh rate, and have the glasses cover alternate eyes at the same refresh rate. A signal is needed to ensure that the glasses are synchronised with the display, but then the user simply sees a different half refresh rate image in each eye[7].

[7]Hammond 1922

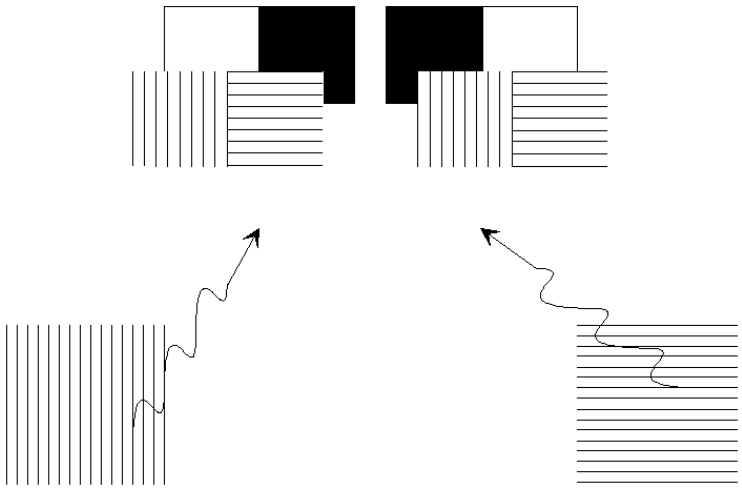


Figure 2.4: Polarisation. Once light has been polarised in one direction (in this case, horizontally or vertically) it can no longer pass through a polarising filter of the other orientation.

These are highly effective, with the sole difficulty in their use (other than the necessity for a synchronisation signal) being their comparatively high cost.

2.3.2.3 Polarised glasses

Light travels in a two dimensional wave. Normally beams of light from a source are transmitted in all possible planes, but light can be polarised - that is, only the light from a certain plane is allowed past. This means that if two images were shown, each transmitted through orthogonally polarised lenses (i.e. vertical polarisation for one and horizontal for the other), then the wrong images could be filtered out through lightweight and inexpensive polarised glasses.

This gives a high quality stereo image, but does have some disadvantages. In its simplest form it requires two projectors or displays for the two polarisations, and in the case of two projectors it requires a screen which will not scatter the light, as this would damage the polarisation and cause cross talk. Recent advances allow for electronically driven polarisers, however, which will alternate between the two polarisations frame by frame on one display device. Another difficulty is that the polarisation at the eyes must be the same as the polarisations on the display - to the extent that if the head angle were to change then they would no longer be in alignment and crosstalk would occur, although this problem can be reduced by using circular instead of linear polarisation.

2.3.2.4 Pulfrich glasses

Clear on one side and darkened on the other, Pulfrich glasses give a spinning object depth by tricking the eye using the Pulfrich effect. The darkened eye has a very small delay in

receiving the data to the brain, giving a small binocular disparity. In this way an object moving horizontally in one direction will appear closer than an object moving in the opposite direction. This has limited use, as the objects must have a constant horizontal velocity to remain at their correct depth.

2.3.2.5 Chromadepth glasses

Chromadepth is a patented system that uses a prismatic film over the eyes which diffracts colour based on its frequency. Seen through them red will appear in the foreground and blue in the background, with the rest of the spectrum of visible light ranging in between. They give a good three dimensional image, but have the disadvantage that the image must be coloured correctly - this could (depending on the object being viewed) detract from its appearance more than, for example, removing colour altogether and converting to monochrome would.

Chapter 3

Theory

3.1 Alignment

For the purposes of this section the two images of the stereo pair, from the left and right perspectives, will be taken as being two equally dimensioned copies of the same image with the second image being a distorted or degraded copy of the first at a particular offset.

Two images of the same scene can be offset in up to four axes:

X horizontal

Y vertical

Z perpendicular into the image, or zoomed

R rotated about the centre

These can be separated into two groups; Rectangular (X and Y) and Polar (R and Z) - that is to say, a change in rotation or zoom is equivalent to a horizontal or vertical change when the images have been changed from rectangular to polar. It therefore follows that if a

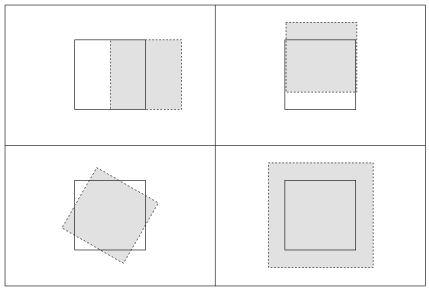


Figure 3.1: Images can be offset in up to four axes: (clockwise) horizontal, vertical, perpendicular or rotational

computer can detect X and Y offsets it can also detect R and Z offsets in the same manner, after converting to polar[1].

As has been discussed previously (see 2.1.4), image alignment can take place in either the spatial or the frequency domains. Detecting offsets in the spatial domain would involve overlaying one image with the other at a variety of potential offsets, keeping a record of the accuracies of the matches at each offset, and then using the offset of the one with the greatest match - that is, to take the best match via a series of comparisons in the spatial domain. This is, however, slow and wasteful of computer resources, making each comparison a $(w \times h)^2$ solution (where w and h are width and height respectively).

The method investigated in this project is to use two dimensional Fourier transforms to allow for a match to be found in the frequency domain, using a technique called Phase Correlation.

This is based on the Fourier Shift Theorem (see Appendix, section 4), which works on the principle that the displacement of objects between the two images is proportional to the phase shift[2].

Two dimensional Fourier transforms are performed on both images, giving a complex equation for each (G_A and G_B - see Fig. 3.2). Taking the complex conjugate of the second image gives $-\phi_B$, and multiplying G_A by G_B^* gives $r_{AB^*}\, e^{(i\,\phi_{AB^*})}$, where $\phi_{AB^*} = \phi_A + \phi_{B^*} = \phi_A - \phi_B$. This is then normalised by dividing by its modulus (i.e. r_{AB^*}). The resulting unit vector $e^{i(\phi_A - \phi_B)}$ gives the phase difference. The inverse Fourier transform is then taken to leave a two dimensional 'image' with a series of peaks. The magnitude of each peak is proportional to the accuracy of the match between the two images at the offset of the peak - hence, the best alignment will be at the offset of the highest peak.

This gives the best alignment in X and Y for the current R and Z, and if the resulting image is converted to polar coordinates the same process will give the best alignment in R and Z for that X and Y. Repeating the process will continually give better alignments until the two images are exactly overlaid.

Using this method provides highly accurate outputs, where even high frequency movement (i.e. movement of a small object) is recorded with an appropriately high peak. However, while this method is very accurate it does present two main problems.

The first is that when looking at a repetitive image (e.g. railings), a peak would be produced for each match (i.e. each railing would match to every other railing as well as to itself); this would, however, be a problem for any method used.

The second is that the edge of the image, where items and therefore Fourier waves are abruptly cut short, can produce edge effects in the Fourier transforms. in full Fourier transforms these do not apply, but in discrete Fourier transforms (DFTs) they can have a significant effect on the transform output. The edge effects appear as lines in the output from the phase correlation, making it difficult for the computer to differentiate between edge

[1] De Castro and Morandi 1987; McGuire 1998, 2001*a,b*

[2] Xie, Hicks, Keller, Huang and Kreinovich 2000; Marsh 2001

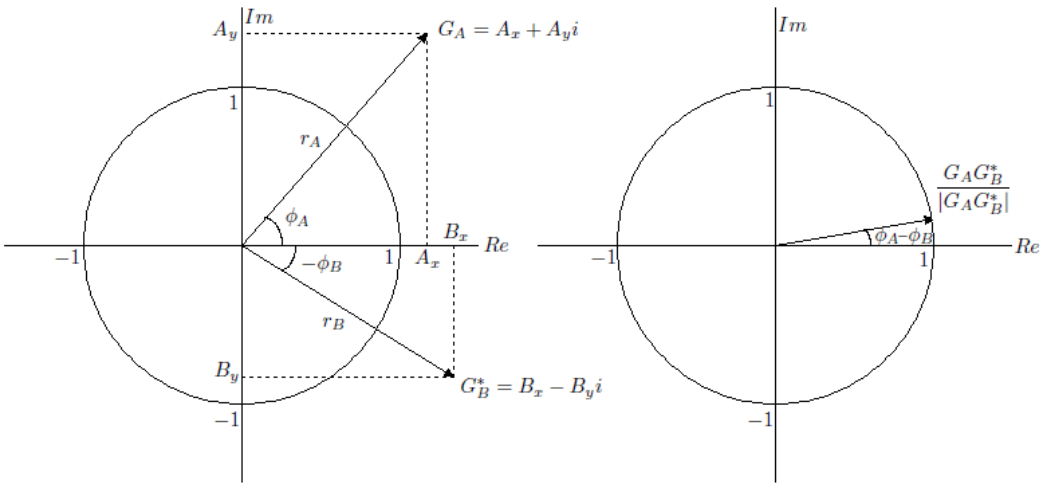


Figure 3.2: Argand diagram demonstrating the Phase correlation technique. Multiplying G_A by G_B^* and normalizing gives the phase difference.

effects and genuine motion peaks. For the most part the effect is sufficiently minimal so as to be ignored; in certain cases, however (notably when comparing polar images in the R,Z comparison) it becomes more prominent.

There are several methods for minimising the edge effects. One simple technique is to add high frequency noise over the image; this breaks up the edge and so reduces the edge effects, but it is difficult to ascertain an exact amount of noise to add to an arbitrary size of image to sufficiently reduce the edge effects while retaining enough of the image to provide a match.

The solution is to remove the edges using a Window Function. One option for this is to remove the edges by fading all of the sides to black in the edge pixels. This works well and removes all edge effects, but is sufficiently processor intensive to render it impractical for the purposes of this project.

As faded edges were processor intensive it was decided to produce a window that while it removed the current edges would itself have hard edges, making its creation faster. As a rectangular window function would cause the same edge effects as before (the new image would have been merely a zoomed portion of the old), a circular window was used - this removes all of the straight lines around the edge of the image, thus preventing them from interfering with the Fourier transforms. While some edge effects are still noticeable in the resulting output from the Fourier transforms these are less intrusive and have a greatly reduced effect on the phase correlation.

Phase Correlation therefore allows the computer to calculate the offset of two misaligned images in all of X, Y, Z and R axes. The final difficulty is to convert the phase correlation offsets into actual quantities. For X and Y this is not a problem - the output from the phase correlation maps exactly onto the input images, and a pixel change in X or Y on the output

is the equivalent of a pixel change in X or Y between the images. In the case of R and Z, however, the output offsets must be mapped into actual rotations and zooms. Rotation can be calculated from the fact that the polar image is a full circle across its height, so that a full range of offsets would cover 360 degrees. The zoom is a more difficult problem to map - but it can also for the most part be ignored, due to the fact that any change in zoom in proportion to the distance of the objects photographed from the camera are likely to be minimal (the change in Z offset would have to be sufficiently large so as to render objects visible in one image out of sight in the other), while any change whatsoever in the other three axes would quickly cause a breakdown in the stereo effect. Removing the calculation of Z from the program also meant that it was only necessary to calculate three of the four axes, allowing for an increase in the speed of program execution.

While the alignment process used is both accurate and considerably faster than many other methods, it is still too slow to be used on the fly to align large images - such as those taken by a modern digital camera - as the Fourier transforms take longer to calculate. This is solved by reducing the size of the image presented to the Fourier transforms and to the phase correlation. A smaller image will complete Fourier transforms faster than a large one, while still giving accurate offsets - assuming enough of the major detail of the image is still discernible. The rotation will remain identical, and a half width image aligned horizontally and vertically and accurate to within a pixel will still resize to give a full sized image accurate to within 2 pixels while making a quarter reduction in the time taken to compute the Fourier transforms. Resizing also reduces another difficulty, which is that the Fourier transforms are only able to be taken on square images with a side length that is a power of two. The images therefore also have their aspect ratio changed and the resized image is ensured to have a side length of 2^n.

Several minor modifications have been made to this basic method to increase quality and efficiency, or to solve a problem. One problem encountered is that when iterating over linear and rotational offsets (i.e. rotating the image, then moving it horizontally and vertically, and then repeating until an accurate match is found) the edges of the image are damaged. This is mainly because there are two options when rotating or shifting linearly as to the edges: either newly unfilled sections are left blank and sections now out of frame are cropped, or the new sections are filled with the cropped sections (i.e. the image wraps around the frame). Whichever is selected, a continuous change alternation of rotating and shifting will lead to much of the edges being cut or filled with non-matching data. Either will shrink the image size and reduce the accuracy of the Phase Correlation, leading to inaccurate matches or more iterations than would be necessary in order to find a perfect match.

One solution to this problem is to keep a master copy of the image being modified, store the accumulated offsets and rotations mathematically (converting rotations not about the centre of the master image into corresponding rotations and offsets). This allows the image presented to the phase correlation to be recomputed when needed, removing the accumulation of edge changes and keeping the image quality as high as the original. There

is a slight reduction in efficiency as the image must be manipulated afresh for each phase correlation, but this is counteracted by the reduced number of phase correlation iterations required to match the images.

3.2 Compression

A stereo pair comprises of two images taken at a slight horizontal offset. They are therefore two images of the same objects, and should therefore be identical save for the horizontal offset of each object between the two images, which is relative to its depth.

As we have two near identical images, we must be storing the same data in both at some offset. If, then, we could store just the changes between the images then we would be able to reduce the size of the stereo pair from image+image to image+difference.

Several methods were attempted when searching for a suitable compression technique for stereo images.

3.2.1 Standard image compression

As both halves of the stereo pair are images, the first attempt at a method of compression was to use standard image compression techniques (see 2.2.2) in an attempt to reduce file size, using a similar method to the JPS format (see 2.2.1). PNG and JPG were tried, and the compression did reduce file size - but, as this does not take into account the differences between the two images, it reduced only to the extent that any two potentially dissimilar images would have had a reduction in file size.

3.2.2 Only store partial FFT

The Fourier transform contains a large quantity of data, split into varying frequencies. If the high or low frequencies were removed from one image, the result would when saved as an FFT image take up less space as the physical dimensions of the image would have reduced when the data was cut. This worked on the hypothesis that the brain would fill in the missing detail from the other image in the stereo pair.

This resulted in reduced file sizes, but the images output at the end had lost a large amount of quality and were noticeably different from their originals. Removing high frequencies quickly blurred the images, and similarly removing low frequencies removed all but the edges. The reduction in image data is so great for a small reduction in Fourier size that the the returned image is unusable in a stereo pair.

3.2.3 Phase correlation peaks

The output from the phase correlation gives the offsets for each point in the image for any frequency. It should therefore be possible, in theory, to save this offset data and reverse the

phase correlation to recreate the image.

Unfortunately the reversal is mathematically impossible due to the normalisation (see 3.1), and if normalization is removed then the phase correlation output is no longer just a product of the phase difference and ceases to contain definite peaks, rendering the offset accuracy unusable.

3.2.4 Manual pixel comparisons

As previously stated, the two images should contain the same object data with a different horizontal offset for each object between the two images. It could therefore be hypothesised that for every pixel in one image there should be a corresponding pixel in the same horizontal line in the other image. The flaw in this hypothesis is that sections of object invisible from one eye are visible from the other, but in the majority of cases the individual pixels of these sections will be similar to the remainder of the object, at a pixel level.

Each pixel can then be tested against every other pixel in the same horizontal line in the other image, while recording the offset which gave the best match. When the image is to be decompressed, the stored offsets give the pixels from the first image which are then used to reconstruct the second.

This was the first method attempted that gave a high quality output image for a high reduction in file size, but was disadvantaged by being processor intensive and slow, as $w \times w \times h$ calculations needed to be made. It also failed to make good use of the stereo effect - the method could be used between any two images with a similar pallet, as the order in which the pixels are checked ceases to be important.

An attempt was made to extend this method by testing blocks of varying size against other blocks in that row[3] - this did decrease time by reducing the number of matches, but each match required more calculations, causing the total time to remain effectively constant. Another downside was that the increase in block size reduced the resolution of the compressed image, which caused edge blurring where the blocks overlapped the side of an object.

3.2.5 Potential overlays

Following on from the pixel comparison method, each pixel should have a match at some offset in that horizontal line. If, then, a series of images were created to catch matches within a certain threshold for every possible horizontal overlay of the entire image (i.e. for each possible horizontal offset an overlay map would be produced that would mark on it pixels where their difference between the two images (at that offset and that pixel) were below a preset threshold), those overlays could then be merged to produce the compressed image.

[3] Fisher 1999; Russ 2005

While this method worked it had two difficulties. The first was that it was also very processor intensive (to a comparable extent as the single pixel comparisons above, if not more so), but a more pressing problem was the level at which the threshold was set - i.e. the level of difference between the pixels of the two images below which they could be taken to be identical. If too low then pixels would find false positive matches before reaching the correct match, greatly reducing the final image quality. If too high then pixels may not find a match at all, leaving holes in the final image. This second problem was not applicable to the pixel comparison method as that matched to the pixel that had the best match - not to the first pixel that had a match within a threshold.

In theory, each pixel in one image should have effectively an exact match in the other - meaning that a high threshold could be used rather than resorting to a more processor intensive best matching scheme[4]. The main reasons why a pixel may not have an exact match are twofold. First, a section of an object may only be visible from the angle of one image and not from the other. With distance this effect is lessened, however, so that only close up images should contain differences in how the object looks. Secondly, and more importantly, the two images may have been taken asynchronously. While the scene remains nearly identical, the light levels may have altered. This would marginally change the brightness and contrast of the image which, while being negligible to the human eye, would alter the pixel values. Asynchronous images may also contain other temporal differences - such as trees moving slightly in the wind, or people walking.

3.2.6 Final solution

The final solution decided upon was to take elements from the phase correlation method and the single pixel matching method above, using the fact that the only changes between a perfect stereo pair should be offsets in the horizontal axis. While there would be other changes in real images, this would be detracting from the stereo effect and it could be hypothesised that these could be ignored without great image loss.

The phase correlation gives a map showing the relative amount of image that matches at each possible value of horizontal or vertical offset. If the values are summed along the vertical axis the result is a single horizontal line, where the values of each point on the line gives the proportion of the image that will match at that horizontal offset. The difficulty is that while this tells us how much of the image is offset by each amount, it does not say which sections of the image have which offsets.

The horizontal values can then be ordered from the highest match to the lowest. A pixel is most likely to find a perfect match on the same horizontal axis in the other image at one of the higher peaks on this new, horizontal phase correlation map. The number of points tested between the images can then be limited, as the majority of pixels will match against a pixel at one of the higher offsets.

[4] Marsh 2001

A threshold T is set for the number of pixels against which each pixel is to be checked, and the offset of the best match from those T pixels is noted for the compression output. This ensures that if a perfect match isn't found, it's likely that a reasonable match has been (as opposed to the potential overlays method, where no match gave a blank pixel). By limiting the number of pixels tested the number of calculations needed reduces from $w \times w \times h$ to $T \times w \times h$ - where T is typically far lower than w while still producing a decompressed image that is difficult to differentiate from the original. Thus, before compression begins, the user can set the maximum time taken by changing the height of the threshold; where a low T is fast but inaccurate, and a high T is slow but accurate. In many cases the number of high peaks is minimal (in theory there should be only one peak for each moved object in the image - in practice, however, there is a lot of noise), which allows a far lower threshold to be used without adversely affecting the image quality.

This produces an original left image and a set of the offsets for each pixel of the right image - in this experiment the offsets were saved as an image, where each pixel's offset was represented as a greyscale value ranging from 0 for a -w/2 offset to w for a w/2 offset. Saving the two images (original right and the greyscale offsets) showed that a substantial reduction in file size had been achieved.

This was further reduced by splitting the greyscale offsets image into two. As there was a defined threshold T for the number of pixel offsets checked, there must be only T possible options for each offset. This allows for the offsets to be stored as a pallet of T colours, where the value of each pixel was the key to a 1xT map to give offsets. Storing the map (a 1xT image with the original offset values) and the offsets (in the reduced, pallated form) allowed the image size to lower far below that of the uncompressed image.

An advantage of this format is that only one image is ever altered. The first image will always remain identical to the original, because it has not been changed - this means that it can be easily viewed by a standard two dimensional display, which allows for backwards compatibility and means that all that is being stored or transmitted is effectively the image and the 'added depth' as a separate file.

The stereo pairs can easily be returned from their compressed state to the original image. For each pixel at (X,Y) in the new Right image the program takes the pixel from the old Left image at the same Y coordinate and at the X coordinate specified by the map, at the point in the map specified by the offset pixels (see Fig. 3.3).

3.3 Beam splitter

As the experimentation to ascertain the quality of the alignment and compression methods required perfectly aligned images to compare against, a secondary objective to the project was to investigate stereo beam splitters (devices which allow one camera to take two stereo images, through splitting the single image taken into two and shifting each left or right to

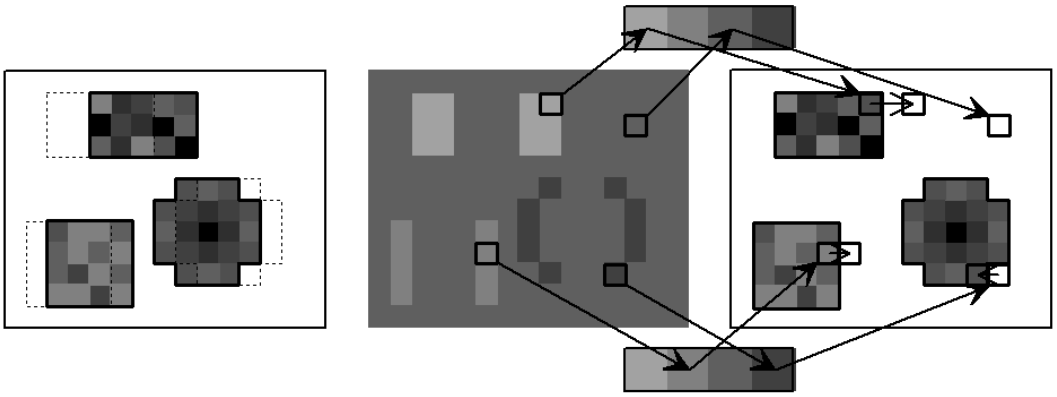


Figure 3.3: Final compression method - the stereo pair is stored as the base image plus a map of changes plus a map of offsets for those changes.

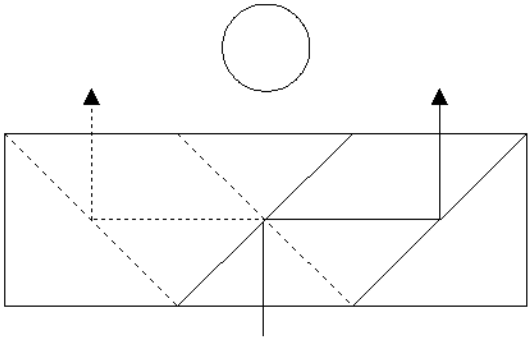


Figure 3.4: Beam splitter (first design), top view. The view from the rear aperture is the scene as viewed from the left and the right apertures, displayed top:bottom.

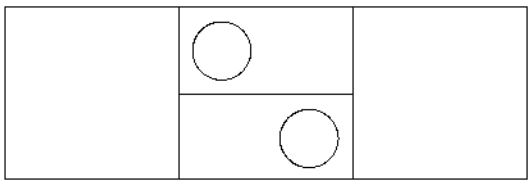


Figure 3.5: Beam splitter (first design), intended rear view. It was believed that both the top and bottom image would contain the whole scene in front of the beam splitter, from the left and right perspectives.

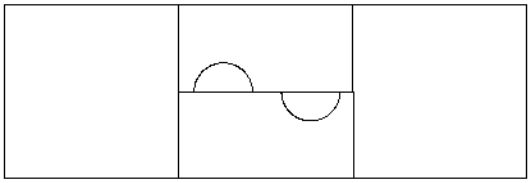


Figure 3.6: Beam splitter (first design), actual rear view. The actual view from the rear aperture was the top half of one image and the lower half of the other, due to an unforeseen vertical offset.

provide an horizontal offset[5]. This was a problem that has been solved previously[6], and so the difficulty in this situation was to decide on a solution that would be easy to produce, in order to create photographs that could be used in comparison with artificially aligned images.

To combine two images onto one camera it is necessary to channel the light from the left image and the right image each fully onto one half of the central picture, centred in front of the camera lens so that the camera can view both as separate images. The split in the picture can be top:bottom or side:side; in the first design top:bottom was arbitrarily chosen. Top:bottom gives panoramic scaled images when taken in a standard landscape orientation, while side:side would give portrait scaled images. Lenses and mirrors can each be used to channel light, and it was first decided to use mirrors to channel the images. This led to the development of the device shown in Figs. 3.4 and 3.5, where two pairs of parallel plane mirrors set at 45° to the viewer and scene are used to view the image from the left and the right, working on the principle that the vertical offset would be negligible with respect to the large field of view. This channelled the light from each front opening into one of the two halves of the camera lens, and ensured that the horizontal offset remained constant by centering the view for the camera along the central axis of the mirrors.

Although the beam splitter did function in the respect that two images were combined from left and right into one split picture, the method did have one major flaw. This was that the field of view from each aperture was curtailed such that the overlap of visible objects

[5]Linssen 1952

[6]Inaba 1998; Haeeis 1874; Tuttle 1946; Maoneille 1946; Warmisham 1934, among many others

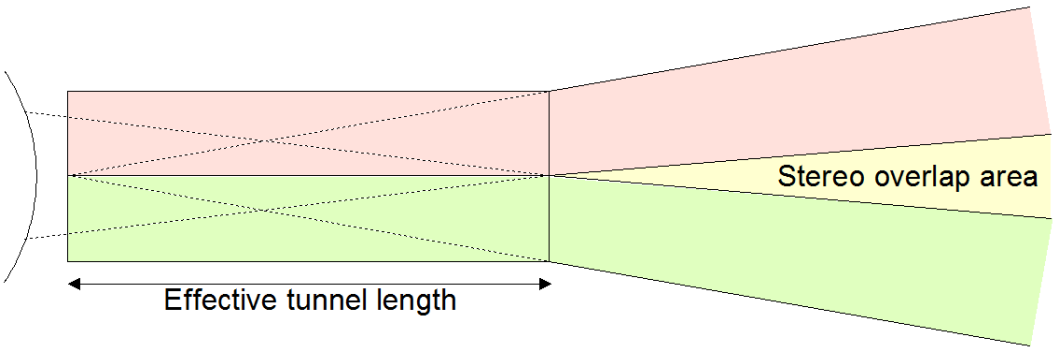


Figure 3.7: 'Shoebox' or Tunnel effect inherent in the beam splitter. When viewed from the rear, the area of overlap between the two images seen is slight due to the length of the tunnel.

was minimal at best, even when viewing at a distance (see Fig. 3.6).

This was discovered to be due to a 'shoebox' or tunnel effect (see Fig. 3.7), where the image being seen reflected through the mirrors was effectively being seen from the end of a 75mm rectangular tube. This allowed each half to see the scene that was immediately in front of it, but made it unable to see anything more than a couple of degrees out from the perpendicular, cutting out the majority of peripheral vision and limiting the angle of view down to a few degrees (see Fig. 3.7). This meant that while the camera was indeed presented with two images with the correct horizontal offset, the two images were also given an undesirable vertical offset. This meant that although the apertures from which the two views were taken were only 10mm apart vertically the two views shown through them failed to overlap significantly even over great distance, rendering them useless.

Attempts were made to rectify the problem, by tilting one pair of mirrors in order to angle one viewed image into line with the other. This, however, introduced a twist into the angled image. Although it could be rectified using the alignment program designed for asynchronous images, that would prevent it from being used for comparison to the aligned images.

To test the stereo splitter a light source was channelled into the camera aperture. This then projected the light via the two outer apertures onto a screen, producing a pair of rectangles corresponding to the centres of the stereo images that would be seen by the camera (see Fig. 4.5). These were then manipulated by attempting to adjust the angles of the mirrors. The change in the projected rectangles was observed as the mirror angle was changed, and it was discovered that the projected beam became distorted as the mirror was adjusted. The distortion worsened as the distance of the beam splitter from the screen increased; as did the vertical offset of the beam shown through the adjusted mirror such that the splitter would only be functional at one specific distance.

Another option would have been to place lenses on the output apertures of the beam

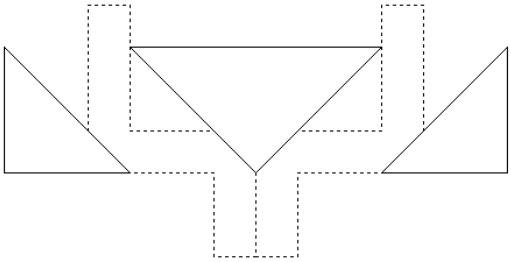


Figure 3.8: Beam splitter (second design), top view. The view from the rear aperture is the scene as viewed from the left and the right apertures, displayed side:side.

splitter. This would give a wider (or, in this case, higher and deeper top and bottom) field of view for each image and greatly increased the overlap between them. The difficulty, however, was that the lenses would have introduced distortions and focussing problems into the images taken. This could in itself have damaged the stereo effect and would also have made the images produced unable to be used for comparison with the computer aligned images.

It was decided to recreate the beam splitter to a slightly different design (see Fig. 3.8). This new design could be created in one of two ways - either channelling the light using mirrors or using prisms. Using prisms would give a more versatile device which would have a wider availability of modification by changing the prism mountings. Using mirrors would give a cheaper solution which would be faster and easier to make, and as it is for a prototype device,it was decided to go with mirrors.

Using mirrors did not preclude the changeable distances from the design, but this would have been simpler when using prisms as the mount design could have the alterations incorporated without introducing excess complexity, while the mirror mountings could be more simply produced from a single block.

The change in design from the original beam splitter was primarily to change from a top:bottom design to a side:side one. This would allow for any offset problems to be horizontal - merely increasing the stereo effect, if having a noticeable effect at all. As this design was finalised and fabricated late in the course of the project it was only given a preliminary testing, using the same method as for the first beam splitter (a light source projecting the output rectangles onto a screen). The results from this experiment (Fig. 4.7)) showed that the rectangles projected were at the correct vertical offsets, regardless of the distance between the screen and the beam splitter.

3.3.1 Synchronised camera stand

Creating a stand to hold two synchronised cameras was also attempted. This problem can be modelled as a beam, with one fixed support in the centre (the camera tripod to which it is attached) and two equal point loads equidistant from that centre (the two cameras)

- which simplifies to two fully fixed cantilevers. While the optimum distance is 70mm[7], it would be preferable for the test assembly to be capable of changing the distance between the cameras, in case this was required later. The stand was therefore made up of sliding blocks running along two horizontal lengths of stud bar, held in position by nuts. Two bars were used instead of one in order to reduce twisting along the length, and each block could be attached from either top or base to a camera or tripod. This modularity and versatility of design allowed it to double as stand for the beam splitter, allowing the camera and beam splitter to be attached to the same mount with a variable offset between them.

[7] Jeong and Bjelkhagen 1992

Chapter 4

Experimentation

This project has covered the designing and programming of a system to align two images to create a stereo pair, and a further system to compress a stereo pair into reduced space. The results of those systems on test images were good; the alignments appeared accurate, and the decompressed images appeared to have lost little or no information when compared to the original input images. The best method, however, for ascertaining whether a stereo pair is accurate or not is to have a human rate it. It was therefore decided to run a series of experiments where volunteers were asked to rate a series of stereo pairs, to see whether the realigned images or the compressed images rated the same or worse than ones taken using a different method (such as with a beam splitter).

To minimise discrepancies between readings all volunteers saw all stereo images using the same three dimensional display - a Dimension Technologies Inc. DTI 2015XLS Virtual Window. Each volunteer saw all stereo images in one sitting, to prevent changes due to light levels. The order in which the stereo images were shown was randomised by a computer, to prevent errors due to ratings changing over the course of the test. A sample randomised order is given in the appendix, in section 2.3.

Two modifications were potentially being made to the images. The first was aligning, the second compressing. Alignment could be made using a proportional sized image to reduce execution time. Compression could be calculated by matching only a proportion of the pixels in the same horizontal bar. Each stereo pair in the test then had an Alignment value and a Compression value. The Alignment value was as a percentage, where 100% corresponds to the actual image size being aligned. Originally it was planned for 0 alignment to refer to an image requiring no alignment, taken through a beam splitter or similar (and when taken using such a method this would be the only option); however, images from a beam splitter were unavailable at the time of testing and so 0 instead refers to the images as taken with no attempt to prepare them for stereo viewing. The Compression value was again as a percentage, where 100% corresponds to the entirety of each horizontal line being tested against each pixel. 0 means that no compression was performed on the image; this

allows for a base quality against which the compression values can be tested. Testing at five values of compression (0 (i.e. uncompressed), 2%, 10%, 50% and 100%) and four values of proportional image size for alignment (0%, 10%, 50% and 100%) gives between 4 and 12 possible options for each image.

The alignment method works best when there are several objects that provide good, clear, prominent outlines which can be easily matched between the two images. It could then be supposed that it will be less likely to work as accurately on 'noisier' images (i.e. images where there are many varied objects).

The compression method works best when there are many comparable pixel values in each line, but also where there are few and definite peaks returned from the phase correlation. The first of these would imply that noisier images are more likely to contain the correct match of colour, but that the second would prefer a cleaner image with few, clear cut objects (similarly to the alignment method).

As the success of both methods may be dependent on the subject matter concerned, it was decided to run the test using a series of images of different types of scene. Each stereo image would therefore be given a Noise level, with the two possible options of High (many and varied objects, mainly taken outside, may include minor temporal changes between images) and Low (few, relatively clean cut objects set in a still life scene, mainly taken indoors, with no temporal changes (as all objects are static and not subjected to external influences)).

This gives 40 possible combinations of Alignment, Compression and Noise, based on a minimum of two different base images. In the experiment four base images were instead used, with two of each type, to allow for differentiation between images (see Appendix, section 2.2). Assuming that the time taken to assess the quality of a three dimensional image is between 15 and 30 seconds, that gives an experimentation time of between 10 and 20 minutes. As this is a long time for a volunteer to retain concentration each test is split into two, such that a pair of volunteers will cover all possible combinations of the 4 images. The number of images shown to each person was later increased to 50, allowing for 10 repeated images. This was to check that the results were accurate and not random. To reduce the time taken to align and compress the images, and to make all image sizes the same, all images were resized to the maximum resolution supported by the three dimensional display prior to use.

It is to be expected that the general public is unused to using three dimensional displays, and there is therefore a possibility that being unused to the screen may affect their results. In order to counteract this each volunteer was instructed as to the basic workings of the screen before the experimentation began. This consisted of an explanation on how the image was transmitted and warnings about keeping their eyes within the correct positions to see the stereo effect. To allow them to find the correct position they were presented with a test screen. This consisted of two dissimilar images (in the case used, one image of red vertical lines marked 'LEFT' and one of blue horizontal lines marked 'RIGHT'), where one image

was transmitted to each eye. Viewing both simultaneously (one to each eye) would have been an unpleasant experience to the viewer as the pair would be impossible to fuse; the volunteer therefore was instructed to close each eye in turn and to align the other such that they could only see the correct image with no crosstalk or ghosting from the other. Once both eyes were independently aligned correctly then it should be possible to see the test stereo pairs without great difficulty.

It is difficult to quantify a qualitative measurement such as 'quality'. In order to enable them to do this, the volunteers were asked to rate the quality of each image on a scale of 0 to 10 and were each given a set of criteria by which to judge the quality on that scale:

Quality	Criteria
10	The stereo effect is perfect. The stereo pair would be indistinguishable from seeing the real scene.
9	
8	The stereo effect is very good, verging on perfect. There is good depth perception throughout the entire image.
7	
6	The stereo image is good. The majority of the image fuses to create a reasonable stereo image, although this is let down by a few objects that fail to fuse correctly.
5	
4	While the stereo image is reasonable, there are large parts which fail to produce a coherent three dimensional image.
3	
2	The stereo image is bad. While some objects are mainly fused into three dimensions, the majority fails to combine and there is large quantities of ghosting (crosstalk)
1	
0	The stereo pair is impossible to view. There is no fusion, and it is unpleasant to see.

4.1 Experimental results

Compression width	% file size of compressed image		
	Min	Avg	Max
2%	17.67	21.35	24.24
10%	29.66	32.72	35.73
50%	61.60	66.58	72.00
100%	68.57	73.70	79.29

Table 4.1: Table of compressed file sizes at a range of compressions, showing maximum and minimum values from test images

Image no.	Noise	Alignment	Compression	Average result	Image no.	Noise	Alignment	Compression	Average result
1	High	0	0	2.8	3	Low	0	0	2.6
1	High	0	2%	1.1	3	Low	0	2%	1.1
1	High	0	10%	1.8	3	Low	0	10%	1.6
1	High	0	50%	2.0	3	Low	0	50%	1.9
1	High	0	100%	2.0	3	Low	0	100%	2.6
1	High	10%	0	4.9	3	Low	10%	0	6.4
1	High	10%	2%	4.2	3	Low	10%	2%	4.4
1	High	10%	10%	4.2	3	Low	10%	10%	4.9
1	High	10%	50%	5.0	3	Low	10%	50%	6.2
1	High	10%	100%	4.5	3	Low	10%	100%	6.7
1	High	50%	0	5.4	3	Low	50%	0	3.5
1	High	50%	2%	7.5	3	Low	50%	2%	2.4
1	High	50%	10%	7.3	3	Low	50%	10%	4.0
1	High	50%	50%	6.6	3	Low	50%	50%	4.7
1	High	50%	100%	8.2	3	Low	50%	100%	4.7
1	High	100%	0	5.6	3	Low	100%	0	6.5
1	High	100%	2%	5.0	3	Low	100%	2%	6.8
1	High	100%	10%	6.6	3	Low	100%	10%	6.6
1	High	100%	50%	7.4	3	Low	100%	50%	6.1
1	High	100%	100%	6.0	3	Low	100%	100%	7.6
2	High	0	0	5.3	4	Low	0	0	2.3
2	High	0	2%	4.5	4	Low	0	2%	0.4
2	High	0	10%	4.4	4	Low	0	10%	1.2
2	High	0	50%	4.8	4	Low	0	50%	1.2
2	High	0	100%	4.5	4	Low	0	100%	1.4
2	High	10%	0	6.6	4	Low	10%	0	4.6
2	High	10%	2%	5.9	4	Low	10%	2%	4.1
2	High	10%	10%	6.8	4	Low	10%	10%	5.6
2	High	10%	50%	5.6	4	Low	10%	50%	3.8
2	High	10%	100%	7.4	4	Low	10%	100%	4.3
2	High	50%	0	6.9	4	Low	50%	0	3.8
2	High	50%	2%	7.5	4	Low	50%	2%	4.9
2	High	50%	10%	7.4	4	Low	50%	10%	6.8
2	High	50%	50%	7.5	4	Low	50%	50%	3.7
2	High	50%	100%	8.5	4	Low	50%	100%	4.9
2	High	100%	0	8.0	4	Low	100%	0	4.2
2	High	100%	2%	7.8	4	Low	100%	2%	3.5
2	High	100%	10%	7.8	4	Low	100%	10%	3.4
2	High	100%	50%	8.0	4	Low	100%	50%	3.8
2	High	100%	100%	8.4	4	Low	100%	100%	4.6

Table 4.2: Summary table of results for all image types, showing the average quality level from all volunteers

Image no.	Variation of repeated images		
	Min	Average	Max
1	0.25	0.83	1.22
2	0.12	1.10	1.90
3	0.00	1.06	1.55
4	0.26	0.66	1.67
Total	0.00	0.91	1.90

Table 4.3: Table of variation between repeated images from each base image, showing maximum and minimum values

Each group of results, making up half of a full set, was normalised individually prior to being combined with the other results. This allowed for variation in the volunteer's perception of quality - if they consistently marked images either poorly, indifferently or well but within a narrow range then their results were stretched to allow for them to be compared equally against the other results.

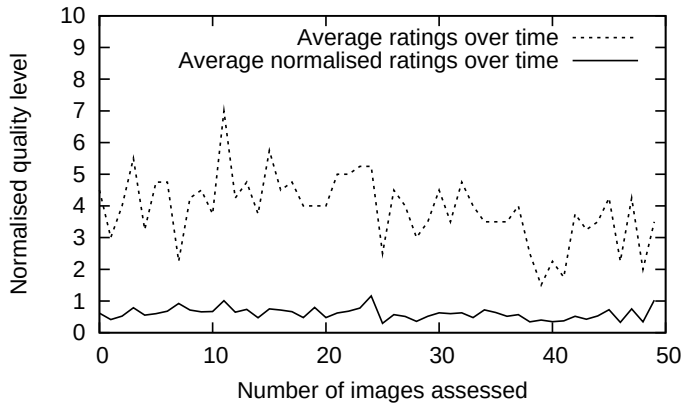


Figure 4.1: Average ratings over time, and normalised average ratings over time (to remove statistical trends from the random ordering)

4.1.1 Repeated images

The repeated images allow for an error margin to be ascertained, as they depict the variation between a volunteer's own results. As can be seen from table 4.1 the average variation given by the repeated images is less than one step, with the maximum remaining under two - i.e. the margin of error is less than ± 1.

4.1.2 Changes over time

One aspect of the experiment which could adversely affect the results is the length of time over which the volunteers are expected to view the images. To become acquainted with

the apparatus and to see a full set of half of the available images takes each volunteer approximately between 15 and 30 minutes. Their assessment of image quality may change over time due to a number of factors. At the start they may still be unused to the equipment and have no previous images against which to compare for quality; at the end their attention may be flagging. Fig 4.1 shows the average result given against time, and it would appear to show a slight downwards trend, implying that there is a change over time. The images shown to each volunteer were in a random order, however, so that any image would be unlikely to appear at the same time for multiple people. The line of average results against time can therefore be normalised against the average result for each image, to show the proportional change in readings over time regardless of the order of images. This second line shows that there is very little variation (less than one quality level) with no discernible gradient - i.e. there is little or no change in quality measurement over time, and the trend seen in the original readings was merely brought about by lower quality images being statistically slightly more prevalent later in the ordering.

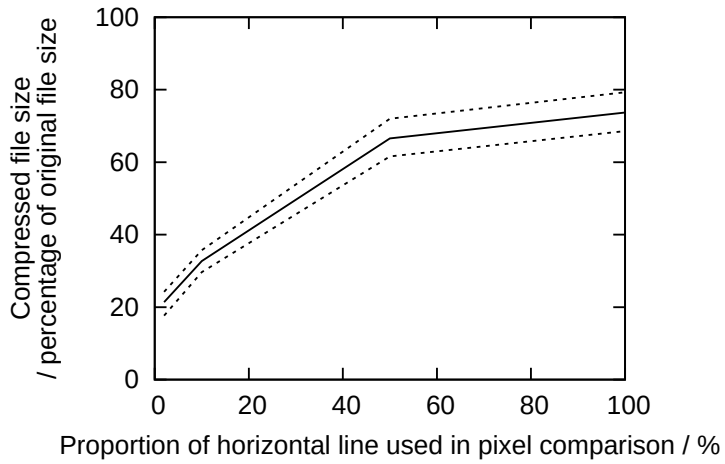


Figure 4.2: Proportional change in file size for various compression levels, shown as a percentage of the original, uncompressed file size. Percentages are taken from the average of all images, showing upper and lower bands.

4.1.3 Compressed file sizes

The compression method used should give a considerable decrease in file size while retaining quality. It is therefore important to verify that a decrease in file size is actually present - this can be seen from Fig. 4.2, where a clear reduction in file size is observable. File sizes are reduced to around 20% at high compressions, and at the highest quality compression level there is still at least a 20% reduction.

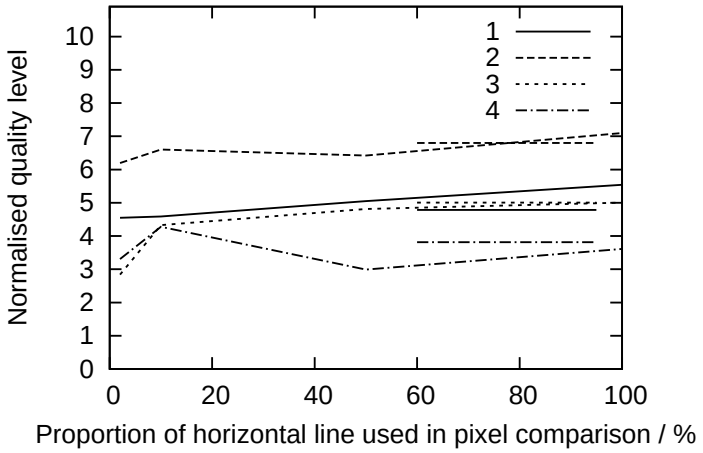


Figure 4.3: Image curves at various compression levels. The horizontal bars refer to the quality of the uncompressed images, for comparison

4.1.4 Compression

On first sight it would appear that very little is shown in Fig 4.3. The four images give greatly varying lines which at most give a slight upward trend, implying that the greater the number of comparisons used in the compression the better the image quality. This, however, is not borne out when compared against the values obtained from the uncompressed images. If the image quality was increasing then the uncompressed images should give the highest quality of all - they do not, however; on occasion the 'perfect', uncompressed image is given a quality comparable to the compression with the minimum number of comparisons.

This can be explained when the error margin from the repeated images is taken into account. The values for all compression levels of each image have a noticeably low range; two are less than one step, and the only image which has a range greater than two steps has its range greatly extended by the point of highest compression.

It can therefore be supposed that to a great extent the variation in the lines is caused more from experimental error than from fact; the trends of Fig. 4.3 show results that are predominantly horizontal, with a slight down turn at the higher compression values. This shows that for the most part the level of compression has no adverse effect on the three dimensional image quality; comparison levels of as low as 10% give very little change in the majority of cases.

There is no evidence that the noise level has any bearing on the quality at different compression rates. While the images have different overall quality levels (i.e. some images are better overall than others), their trend lines keep to a reasonably consistent gradient. This implies that the compression method will work on images regardless of the noise level of their content.

Further graphs are available in the Appendix, in section 2.4.2.

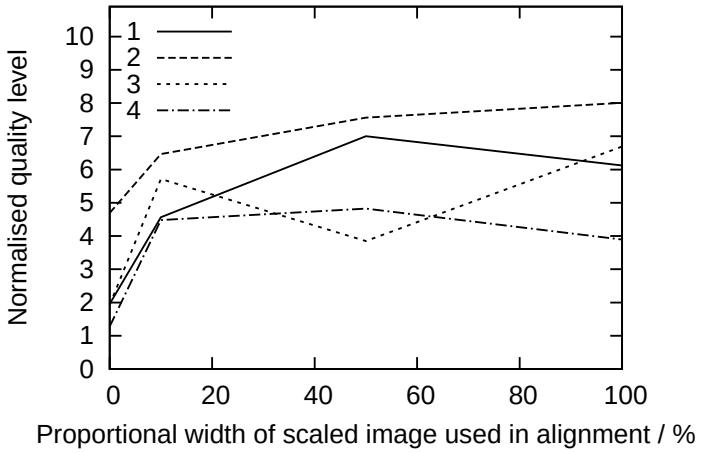


Figure 4.4: Image curves at various alignment levels. Note that 0 refers to the image prior to alignment

4.1.5 Alignment

The curves shown in Fig 4.4 are more complex. There is a definite upward trend at the start, showing that there is a strong improvement in the three dimensional effect from no alignment (the original images, taken asynchonously and by hand) to basic alignment (alignment using a small scale copy of the image to improve speed). For the remainder it was believed that as the scale of the copy used in the alignment process approaches the size of the original image the accuracy of the alignment and so the quality of the image improves. This has been borne out to some degree on the majority of the four curves; that of image 2 shows the expected trace, although the remaining curves do not show a consistent increase in quality. In all three which do not there is one point (i.e. one size of scale image) which changes the curve from its predicted output.

This leads to the hypothesis that while the majority of image sizes align correctly it is possible that one size of a certain image will give a worse than anticipated result. This could be caused by the phase correlation. As the alignment (in all four axes - see 3.1) is based on the relative height of peaks, if two peaks are of a similar height then the one selected as highest may be influenced by the dimensions of the image. This means that some images may give a bad alignment at certain scale image sizes, as can for example be seen in the Image 4 100% alignment (see Appendix, 2.2), where the image was aligned using the original image size but produced an image that was of a similar alignment to the unaligned image. A future project could investigate this to see whether it is possible to ascertain the quality of alignment at a certain scale image size before the alignment has taken place.

An alternative possibility is that the quality does not follow a linear or logarithmic formula dependent on alignment levels. This would, from the theory used to produce the images, appear unlikely - the alignment percentages used in the experiment do not refer

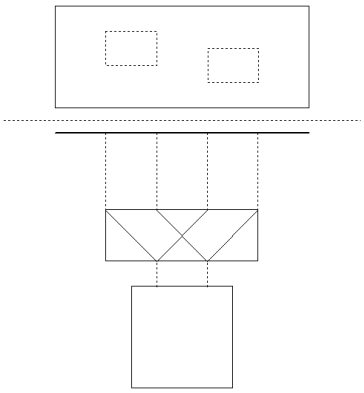


Figure 4.5: The layout for the beam splitter experiment. The light source projects through the splitter, creating two rectangles onto the screen which can be then used for alignment.

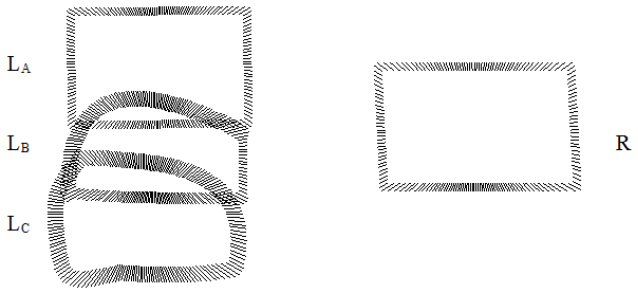


Figure 4.6: The results from the beam splitter experiment, for the first beam splitter. A: Straight projection; L_A and R (100mm). B: Angled mirror to remove vertical offset; L_B and R (100mm). C: Retained same mirror angle as B; L_C and R (150mm).

to the level of alignment used on the image, but merely the size of the scale copy of the image that was used in alignment. In all cases the stereo pair produced are at the optimum alignment for that size of image - the alignment calculated is then scaled back up to the original image. It is still, however, possible for it to not follow a linear or logarithmic formula - further experimentation would be required on a greater range of images to ascertain this.

4.2 Beam splitters

A secondary experiment was also performed to ascertain the extent of the vertical offset of the original beam splitter design (see 4.5), and to see whether it were possible to rectify this by manipulating the mirrors. This was achieved by projecting a non-diffuse light source through the camera aperture of the beam splitter and noting where the projected light fell on a screen erected a set distance away. The experiment was further extended to include the second beam splitter design for comparison.

Fig 4.6 shows the projected rectangles for the original design. At the start of the ex-

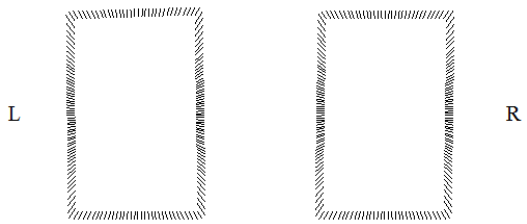


Figure 4.7: The results from the beam splitter experiment, for the second beam splitter. Straight projection; L and R (at both 100mm and 150mm).

periment the screen was placed 100mm away from he light source, with all mirrors in their original positions, giving the projections L_A and R. This shows the high discrepancy between the two projected rectangles - there is a large vertical offset which prevents the use of the beam splitter.

One mirror from the path of the left beam was then adjusted to bring down the left beam in line with the right, giving the projections L_B and R. While these are now aligned vertically, the left projection can be seen to be no longer rectangular - as the mirror has been adjusted the projected image has become twisted and deformed. This has two disadvantages for the beam splitter. The first is that the image perceived by the camera on the left side will no longer be straight. While this could be rectified in post-processing, it would defeat the purpose of the beam splitter - that is, to produce a method whereby a 'perfect' stereo image can be taken with no processing necessary.

The second is that of convergence. Originally all mirrors were only rotated in one direction from the plane normal to the camera view, so that the distance between the camera and the screen, in terms of vertical offset, was largely immaterial. At any distance the centre of the projected image would retain the same offset. Now that the vertical offset is governed by the mirror angle that property changes. Changing the screen distance to 150mm (while keeping the mirror angle constant as at L_B) produced the projections L_C and R. While R has remained constant, L_C has dropped over the distance to a negative vertical offset comparable to the original and has increased in deformity. Changing the mirror angle, therefore, does not give either an effective or consistent solution.

The same experiment was then performed using the second beam splitter design, giving the results shown in Fig 4.7. As can be seen, the images are aligned vertically; changing the screen distance from 100mm to 150mm retained the centres of the projected rectangles in the same places. The second beam splitter design, then, would appear to be the more suitable for its purpose; although further experimentation would be required to ascertain the relative quality of images taken with the beam splitter as compared to those taken by other means.

Chapter 5

Discussion

This project has focussed on various aspects in the field of stereo imaging.

5.1 Alignment

One of the aims of the project was to produce a system which would allow two misaligned images from a stereo pair to be realigned, such that two asynchronous images taken by hand with a minimal amount of aiming could be converted into a usable stereo pair. To a great extent this has been achieved. The code produced is able to align images in three axes (X, Y and R) and, when put into a stereo viewer, these fuse into a stereo image. The experiments performed in section 4, and shown in the appendix at 2.4.1, show the qualities of images as the size of the image scaled down and aligned (see 3.1) changes in proportion to the original image; while there is some evidence from this (see 4.1.5) that some scales of images give unpredictably bad quality alignments, there is still a definite improvement in quality as alignment is introduced.

The system as it currently stands does, however, have flaws. It is unable to adjust for keystoning or alterations in zoom. The calculation of a zoom offset could be added to the code, although there is then the problem of configuration - how great a change in zoom is required for a set value of offset from the phase correlation.

Another difficulty with the current method is that it has no knowledge of the screen depth, relative to other objects. Three dimensional displays have a greatly limited depth of vision - both as only objects that are visible from both eyes should be seen, and because some displays are, in practice, limited to a far greater extent (this has been calculated to be as low as 55mm from the screen plane[1]). The position of objects relative to the screen is set by the offset of that object between the images (see Fig. 1.6). If the object is in exactly the same place on both images then it will appear to the user to be at the same depth as the screen (i.e. sitting in the plane of the screen). A positive horizontal offset will

[1] Holliman 2004

move the object back into the screen (away from the user) and a negative horizontal offset will move the object forwards out of the screen (towards the user). The program aligns the two images with effectively the largest match across the image. In other words, if there is a large foreground object that's the most prominent object to match between images, that will be set to the same position in both images - making the object sit at the distance of the screen. If, on the other hand, the scene comprises several small foreground objects but a very large and consistent background then it is the background that will be set to the same position in each image, while the foreground objects are in front of it. The difficulty then arises when objects come forward out of the screen or go backwards into the screen by too large an amount. If they are too close to the user then he will be forced to go cross eyed to see them; the closer they get the more improbably cross eyed he would need to become in order to fuse the images, which quickly breaks the stereo effect (especially if the objects are not centred in the image). If they are too far from the user then there is a chance that they are impossible to fuse into one stereo image, as the eyes would be required to turn away from themselves, divergent, in order to match the images (this is less likely, however, unless keystoning has been introduced and the images are 'toed-in').

A further problem is when images contain repeated data. If, for example, the image contains a prominent set of railings then the alignment process will be unsure as to which railing in the first image is supposed to match to which railing in the second. This would only be a major problem if the railings were so prominent as to make their alignment more important than any other object in the scene, at which point the program may potentially align the image against the wrong one.

5.2 Compression

The project set out, in part, to produce a method whereby a stereo image pair may be saved in less space than would be taken by two separate images. This, again, has been achieved - stereo pairs can be saved with a significant reduction in overall file size and then be reloaded with little or negligible loss of visible quality. This is borne out in the experiments performed in section 4, shown in the appendix at 2.4.2, which show the qualities of images as the compression rate changes (where 100% corresponds to minimum compression - see 3.2.6). This gives results consistent with the compression requirements - i.e. it shows that there is little or no change in discernible quality outside the margin of error for a change in compression, while 4.1.3 shows a consistent decrease in file size.

It is not, however, a perfect method, in that it is lossy - in other words, the image produced after a compression/decompression is not identical to that which was entered, due to the fact that the pixels in the left image may not be an identical set (albeit in a different order) to the pixels in the right image. In most cases, with the threshold for the number of peaks used from the phase correlation set sufficiently low (i.e. checking each pixel against a sufficiently high number of alternatives), any differences between the original image and

Figure 5.1: A stereo pair, pre-compression. Sections of the sky visible in the first image are occluded by trees in the second image. This data is therefore unavailable in the corresponding horizontal lines, leading to compression errors

Figure 5.2: A stereo pair, post-compression. As can be seen on this closeup of the sky, the data missing from the other image has been filled with data that is not wholly accurate, causing noticeable compression errors.

the compressed/decompressed image will be negligible, and unnoticeable to the human eye. As the 'incorrect' information is being taken from a near identical image, the changed pixels are likely to be from a close point on the same object.

The main case where this does not work, however, is where there is an object entirely hidden between the images. In Fig. 5.1, some sections of the sky are only visible above the trees in one image. When compressed there is little or no data for the sky to be recreated from - the data used is therefore very dissimilar to the original data, making noticeable artefacts on the image where differing pixel colours are used (see Fig. 5.2). In the majority of cases, however, this effect is minimal if occurring at all.

The other difficulty with the method selected in particular is that it only tests each pixel against a small subset of the available other pixels - at a set number of offsets. In most cases this will cover all pixels that are similar as the pixel offsets represent the offsets of the major objects in the scene. However, if the scene contains many small objects that have moved then the pixels selected as possible alternatives may not include the optimum match - or,

in the case of a lot of motion with a low number of offsets, it may not include a match that is even from the same object. This can be counteracted by raising the number of offsets tested, but while increasing quality this will raise image sizes after compression and greatly increase the time taken to compress the image by $w \times h$ calculations per additional offset. Most scenes are unlikely to be affected by this, however.

While this Fourier based method of compression gives a reasonably accurate output that can pick up on high frequency movement and is fast enough to be used for compressing images, it is not fast enough to be used in on the fly video compression. There are, however, alternatives which could be investigated in a future project.

Frames in a video change over time, and the problem of compression is near identical to that with stereo images - in that two consecutive frames will usually hold effectively the same data but with certain objects having moved. The way that the compression is achieved, in its simplest form, is to store a complete picture once every n frames, where for the frames in between all that is stored are the changes made in that frame. The changes are calculated using faster but less accurate block comparison techniques, where an area of the first frame is matched to an area of the second, then a subsection of that first frame area is matched to a subsection of the second frame area; continuing until it is matched. This, however, can have problems with its accuracy - for example, when an object moves from one area to another - although usually any flaws are minimal, and the refresh rate of the video ensures that the viewer is unaware of their existence. This same method could be extended to not only match between frames but also to match from one frame to its stereo pair. If the encoding and decoding algorithms could be extended to cover this then it would be able to perform in real time, and the base algorithms for the stereo compression would be already available.

Other elements of this which could be tested include whether the human brain notices two dimensional visual differences more or less than three dimensional object motion - in other words, whether the difference image could be stored as a separate stream which was not required to be updated at as high a refresh rate, thus reducing data transfer. Another option would be to store a difference image once every n frames and to only store the changes to that difference image for the frames in between - this would mean that for a relatively static scene, where only one object was moving, only that object's new three dimensional offset would need storing - the other objects in the scene would remain as they were.

Another possibility is to convert the offsets from the stereo pair into depths and use the data to create a basic three dimensional model of the scene, over which a single amalgamated image could be stretched. The data transmitted would then be one single image plus the vector data to make up the three dimensional model. If the model was kept as a series of simple shapes (cuboids, planes and spheres) then the amount of space taken up to store the model data would be minimal. An added advantage to storing and transmitting the data in this way would be that the user could then move around the model - although there would be discrepancies where sections of the model were invisible from both images, and

the computer generated model would be unlikely to be accurate to great detail (more so as the user moved from the original centre of the images). This could be continued by taking a bank of images, each at a small offset from the previous one, and using them to build up a more accurate three dimensional model. This would have many uses in a large range of fields - from medical images that could be moved around to more realistic computer game models.

As part of this project a program was written to allow for less trained users to input, align, compress and output stereo images (the source code is included in the Appendix, section 3). A future project could extend this to provide a fuller user interface and improve portability. An asynchronous web interface could also be included from a server; this would help to bring alignment for stereo photography to a wider audience.

5.3 Beam splitter

The beam splitter was one part of the project that was unable to be tested as well as the other sections. This was due to the first design having an unexpected vertical offset that prevented the images from overlapping, which in turn meant that test images could not be taken to compare against those taken asynchronously with one camera and aligned on the computer. Experiments were made to ascertain the exact offsets of the two images by channelling light down the two halves of the pair and thus projecting their visible regions onto a screen. This was extended by altering the angles of the mirrors on the device to discover whether it would be possible to modify the design so as to overlap the images. The results from this (see 4.2) were that the two projected rectangles representing the fields of view did not overlap without an angling of the mirror, which in turn induced distortion into the manipulated image. The angling also will only work to align the images to one depth plane; i.e. objects further in front or behind the object focussed upon will still have a vertical offset depending on their depth offset from the point of focus.

A second design was created to rectify the problem; this again used four mirrors but with all four in a full vertical plane to prevent vertical offsets. The second beam splitter was constructed close to the end of the project, precluding the possibility of a large number of experiments. While no test images have be taken, however, it has been tested using the above method (using a projected light source and screen) and has shown two side:side rectangles which remain in relative alignment irrespective of the distance from beam splitter to screen.

Chapter 6

Conclusion

6.1 Alignment

The method of alignment decided upon, using two dimensional Fourier transforms and phase correlation, appears to work very accurately and enables the creation of well fused, headache-free stereo pairs asynchronously with only one camera. It does have some flaws, the main one being that the objects in the stereo image are not necessarily well laid out in terms of depth as the screen position in the depth field is unknown, but that can be rectified by the selection of the scene and can be manually accounted for.

6.2 Compression

The compression method has some drawbacks, predominantly in that it is lossy compression and can give inaccurate results if there are objects in one image not visible from the other. Other than this, however, it gives results with a reasonably fast algorithm, with a variable level of compression, and which are in the majority of cases of negligible difference from the originals.

6.3 Beam splitter

The design first decided upon did not work due to unforeseen optical difficulties that prevented the two images from overlapping. The second design appears to rectify this, but as yet there has been insufficient experimentation to verify this.

Bibliography

Brown, L. G. (1992), 'A survey of image registration techniques', *ACM Computmg Surveys* **24**(4), 325–376.

Cruz-Neira, C., Sandin, D. J. and DeFanti, T. A. (1993), 'Surround-screen projection-based virtual reality: The design and implementation of the CAVE', *International Conference on Computer Graphics and Interactive Techniques - Proceedings of the 20th annual conference on Computer graphics and interactive techniques* pp. 135–142.

De Castro, E. and Morandi, C. (1987), 'Registration of translated and rotated images using Finite Fourier Transforms', *IEEE Transactions on pattern analysis and machine intelligence* **PAMI-9**(5), 700–703.

Diner, D. B. and Fender, D. H. (1988), 'Dependence of Panum's fusional area on local retinal stimulation', *Journal of the Optical Society of America A* **5**, 1163–1169.

Euclid (1945), 'The Optics', *Journal of the Optical Society of America* **35**(2), 357–372.

Fisher, R. B. (1999), 'Temporal stereo vision: A structure from motion perspective', CVonline: The Evolving, Distributed, Non-Proprietary, On-Line Compendium of Computer Vision; University of Edinburgh School of Informatics. Available from: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/CROSSLEY1/research/temporal_stereo.html [accessed December 2007].

Flack, J., Harman, P. V. and Fox, S. (2003), 'Low-bandwidth stereoscopic image encoding and transmission', *Proceedings of SPIE* **5006**, 206–214.

Haeeis, W. (1874), 'Improvement in stereoscopic cameras', Patent, no. 151973.

Haering, N. and da Vitoria Lobo, N. (2001), *Visual Event Detection*.

Hammond, L. (1922), 'Stereoscopic motion-eectuke device [Teleview system]', Patent, no. 1506524.

Holliman, N. S. (2004), 'Method and apparatus for generating a stereoscopic image', Patent, International Publication No. WO 20051060271 A1.

Inaba, M. (1998), 'Stereo camera', Patent, no. 5778268.

Jeong, T. H. and Bjelkhagen, H. I. (1992), *International Symposium on Display Holography*, Society of Photo-optical Instrumentation Engineers.

Lentjes, A. e. (2006), 'Making a 3-D movie on any budget - a step-by-step guide that will take you from 3-D script to 3-D screen', Available from: http://www.the3drevolution.com/3dscreen.html [accessed summer 2007].

Linssen, E. F. (1952), *Stereo-photography in Practice*, Fountain Press.

Maoneille, S. M. (1946), 'Range finder', Patent, no. 2403732.

Marsh, D. (2001), 'Temporal rate conversion', Available from: http://www.microsoft.com/whdc/archive/TempRate.mspx and http://www.microsoft.com/whdc/archive/TempRate1.mspx [accessed December 2007].

McGuire, M. (1998), An image registration technique for recovering rotation, scale and translation parameters, Technical report, NEC Tech Report.

McGuire, M. (2001*a*), 'Image registration method', Presentation. Slides available from: http://www.cs.brown.edu/ morgan/papers /(McGuire98)RecoveringRotationTranslationAndScale.ppt [accessed January 2008].

McGuire, M. (2001*b*), 'Image registration method', Patent, US patent no. 6,266,452.

Purves, D., Augustine, G. J., Fitzpatrick, D., Lawrence, C. K., Lamantia, A.-S. and McNamara, J. O., eds (1996), *Neuroscience*, Sinauer Associates.

Roberts, J. W. and Slattery, O. T. (2000), 'Display characteristics and the impact on usability for stereo', *Stereoscopic displays and virtual reality systems VII* **3957**, 128–137.

Rosamond, M. (2004), Stereo television and data compression, Meng, University of Durham School of Engineering.

Russ, J. (2005), 'The image analysis cookbook 6.0', Available from: http://www.reindeergraphics.com/index.php?option=com_content&task=view&id=173&Itemid=121 [accessed January 2008].

Silverman, R. (1993), 'The stereoscope and photographic depiction in the 19th century', *Technology and Culture* **34**(4 - Special Issue: Biomedical and Behavioral Technology), 729–756.

Tuttle (1946), 'Stereo system', Patent, no. 2396902.

User Manual for the DTI 2015XLS Virtual Window^{TM} (2001), 315, Mt. Read Blvd., Rochester, NY 4611.

Wade, N. J., Ono, H. and Lillakas, L. (2001), 'Leonardo da Vinci's struggles with representations of reality', *Leonardo* **34**, 231–235.

Wann, J. P., Rushton, S. and Mon-Williams, M. (1995), 'Natural problems for stereoscopic depth perception in virtual environments', *Vision Research* **35**(19), 2731–2736.

Warmisham, A. (1934), 'Optical device', Patent, no. 1964968.

Wheatstone, C. (1838), 'Contributions to the Physiology of Vision - Part the First. On some Remarkable, and hitherto Unobserved, Phenomena of Binocular Vision', *Philosophical Transactions of the Royal Society of London* **128**, 371–394.

Xie, H., Hicks, N., Keller, G. R., Huang, H. and Kreinovich, V. (2000), 'Automatic image registration based on a FFT algorithm and IDL/ENVI', *Proceedings of the ICORG-2000 International Conference on Remote Sensing and GIS/GPS* .