AN INVESTIGATION OF POTENTIAL BENEFITS
OF HYPER-STEREOSCOPIC VIDEO
FOR AERIAL SEARCH AND RECONNAISSANCE

by

Kit Man Cheung

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Department of Mechanical and Industrial Engineering
University of Toronto

© Copyright by Kit Man Cheung 2000
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non-exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-50423-9
An Investigation of Potential Benefits of Hyper-stereoscopic Video for Aerial Search and Reconnaissance

by

Kit Man Cheung

A thesis submitted in conformity with the requirement for the degree of Master of Applied Science

Graduate Department of Mechanical and Industrial Engineering
University of Toronto
2000

Abstract

The purpose of this thesis was to evaluate the potential benefits of using a stereoscopic video system during Aerial Search & Rescue (SAR) operations. The complete project consisted of two different components: 1) the design and implementation of a prototype system for SAR, and 2) evaluation of theoretical benefits of hyperstereopsis. Two laboratory studies were conducted to evaluate two variables that were critical to operator performance using the prototype.

The first study was designed to evaluate potential performance benefits of hyperstereopsis provided by large viewpoint separation. It was found that individual differences play an important role in defining a possible "optimal" viewpoint separation for the system. The second study was conducted to test the effect of increased viewpoint separation on tolerance of vertical disparity. With increased viewpoint separation, binocular fusion would become less stable and tolerance would be reduced. The results provided guidelines for future implementation.
Acknowledgements

I would like to thank Professor Paul Milgram, Dr. Julius Grodski, Mr. Bart Hall, Dr. Shi Yin and Mr. David Drascic for their invaluable assistance through the duration of the project. The thesis was made possible with the generous financial and material (access to information, equipment loans and access to personnel) support of Defence & Civil Institute of Environmental Medicine (DCIEM). Thank you to Rishi Bageshwar for collaborating with me and carrying out experiment 2. I would also like to express thanks to those members of the Canadian Forces who gave their time during the visit to CFB Trenton and CFB Kingston making crucial interviews and testing possible. To my family, I appreciated your support for the past few years. Thank you for being so understanding and patient. To all my friends, both from within the department and from my earlier years, thanks for helping me focus and, at the same time, providing some much needed distraction.
# Table of Contents

**ABSTRACT** ........................................................................................................ II

**ACKNOWLEDGEMENTS** ....................................................................................... III

**TABLE OF CONTENTS** .......................................................................................... IV

**LIST OF FIGURES** ............................................................................................... VI

**LIST OF APPENDICES** ........................................................................................ VII

1 **INTRODUCTION** .............................................................................................. 1

2 **THEORETICAL BACKGROUND** ......................................................................... 4
  2.1 **COMMON STEREOSCOPIC DISPLAY APPLICATIONS** ....................................... 4
  2.2 **AERIAL PHOTOGRAMMETRY** ............................................................................ 5
  2.3 **DEPTH PERCEPTION IN SAR OPERATIONS** ...................................................... 6
      2.3.1 **Environmental Depth Cues** ......................................................................... 8
      2.3.2 **Convergence and Accommodation Cues** ................................................... 9
      2.3.3 **Binocular Disparity** ..................................................................................... 9
      2.3.4 **Hyperstereopsis** ......................................................................................... 12
      2.3.5 **Accommodation-Vergence (AV) Mismatch** ................................................ 14
  2.4 **SAR VISUAL DETECTION** ................................................................................ 15

3 **SYSTEM DESIGN & PROTOTYPE** ...................................................................... 17
  3.1 **DESIGN CONSIDERATIONS** .......................................................................... 17
      3.1.1 **Dual Camera Stereo Alternative** ................................................................. 17
      3.1.2 **Single Camera Motion Stereo Alternative** ............................................... 22
  3.2 **SAR TECHNICIAN INTERVIEW: TRENTON FIELD TRIP** ............................... 24
      3.2.1 **Search Vehicle** .......................................................................................... 24
      3.2.2 **Search Task** ............................................................................................... 25
  3.3 **CONSIDERATION OF PRELIMINARY DESIGN FEASIBILITY** ......................... 28
  3.4 **SELECTED DESIGN** ....................................................................................... 30
      3.4.1 **Design Specifications** .................................................................................. 30
      3.4.2 **Hardware Design** ....................................................................................... 34
      3.4.3 **Software Design** ......................................................................................... 38
      3.4.4 **Tradeoffs** ................................................................................................... 41
  3.5 **PRELIMINARY FIELD EVALUATION: KINGSTON, ONTARIO** ....................... 42
4 EXPERIMENTAL EVALUATION & PERFORMANCE TEST ........................................... 45
   4.1 EXPERIMENT: VISUAL DETECTION USING HYPERSTEREOSCOPIC VIDEO .................. 46
      4.1.1 Hypothesis ........................................................................................................... 46
      4.1.2 Experimental Approach ........................................................................................ 47
      4.1.3 Design of Stimulus ................................................................................................ 49
      4.1.4 Participants ............................................................................................................. 52
      4.1.5 Experimental Design .............................................................................................. 52
      4.1.6 Setup & Procedure .................................................................................................. 53
      4.1.7 Results & Discussions .............................................................................................. 54
   4.2 PERFORMANCE TEST: VERTICAL DISPARITY TOLERANCE .................................... 65
      4.2.1 Objectives ............................................................................................................... 65
      4.2.2 Performance Test Method ....................................................................................... 66
      4.2.3 Stimulus .................................................................................................................. 67
      4.2.4 Participants ............................................................................................................. 68
      4.2.5 Performance Test Design ......................................................................................... 68
      4.2.6 Setup & Procedure .................................................................................................. 69
      4.2.7 Results & Discussions .............................................................................................. 70
5 CONCLUSIONS ............................................................................................................. 73
   5.1 EXPERIMENT 1: HYPERSTEREOPSIS ....................................................................... 73
   5.2 PERFORMANCE STUDY: VERTICAL DISPARITY TOLERANCE .................................... 74
6 RECOMMENDATIONS .................................................................................................... 76
   6.1 RECOMMENDED TECHNICAL ADVANCES ................................................................. 76
   6.2 RECOMMENDED EXPERIMENTAL WORK ................................................................ 77
7 REFERENCES ............................................................................................................... 78

APPENDIX A - EXPERIMENT 1 - SUPPLEMENTARY MATERIAL ........................................ 82
APPENDIX B - EXPERIMENT 1: ANOVA ........................................................................... 85
APPENDIX C - EXPERIMENT 1: ROC CURVES ................................................................. 86
APPENDIX D - EXPERIMENT 2: INTERPOLATION GRAPHS ........................................... 93
List of Figures

FIGURE 1 - THESIS OVERVIEW ........................................................................................................... 3
FIGURE 2 - BINOCULAR DISPARITY ..................................................................................................... 10
FIGURE 3 - DUAL CAMERA STEREO SYSTEM WITH CONVERGING GEOMETRY ................................. 19
FIGURE 4 - TYPICAL OPTICAL FLOW ENCOUNTERED IN DAILY LIFE .................................................. 21
FIGURE 5 - DIAGONAL MOVEMENT OF OBJECTS ON GROUND ACROSS THE SCREEN IN THE PRESENCE OF CROSSWIND ........................................................................................................... 21
FIGURE 6 - RECOMMENDED VISUAL SEARCH PATTERN, IN AIRCRAFT FRAME OF REFERENCE, BASED ON ONLY HEAD AND/OR EYE FIXATIONS OF THE SAR TECH. ................................................................................................. 26
FIGURE 7 - RESULTING VISUAL SWEEP OF TARGET AREA FROM FIG. 6, IN GROUND CO-ORDINATES. ................................. 27
FIGURE 8 - TYPICAL FLIGHT PATTERN OVER SEARCH REGION ................................................................ 28
FIGURE 9 - VIDEO FLOW IN SINGLE CAMERA SETUP ........................................................................... 32
FIGURE 10 - VERTICAL DISPARITY RESULTING FROM LATERAL MOTION OF AIRCRAFT ......................... 33
FIGURE 11 - ELIMINATING VERTICAL DISPARITY BY ROTATION OF CAMERA ABOUT ITS OPTICAL AXIS ................................. 36
FIGURE 12 - SIDEWAYS VIEW VS. STRAIGHT DOWN VIEW ...................................................................... 37
FIGURE 13 - RELATIONSHIP BETWEEN THE MODULES OF THE SOFTWARE .............................................. 40
FIGURE 14 - HYPOTHETICAL DETECTION PERFORMANCE AS A FUNCTION OF HORIZONTAL DISPARITY ................................................................. 47
FIGURE 15 - MODEL USED FOR EXPERIMENT 1. (SAME FIGURE WAS SHOWN TO PARTICIPANTS TO EXPLAIN THEIR TASK. ......................................................................................................................... 51
FIGURE 15B - ACTUAL APPEARANCE OF THE EXPERIMENTAL STIMULUS (IN GREY SCALE). ....................... 51
FIGURE 16 A & B - BASIC DETECTION PERFORMANCE: (A) MEAN PROBABILITY OF HITS, (B) MEAN PROBABILITY OF FALSE ALARMS .................................................................................................................. 55
FIGURE 17 - AVERAGE DETECTION PERFORMANCE (D') FOR ALL 10 PARTICIPANTS ................................. 56
FIGURE 18 - DETECTION PERFORMANCE (D') FOR INDIVIDUAL PARTICIPANTS ...................................................... 57
FIGURE 19 - MEAN D' PERFORMANCE WITH "YES" RESPONSE CUT OFF AT 2 .................................................. 58
FIGURE 20 - INDIVIDUAL SUBJECTS' DETECTION PERFORMANCE (D') WITH CONFIDENCE RATING OF 2 AND ABOVE BEING CONSIDERED AS A "YES" RESPONSE ......................................................................................... 60
FIGURE 21 - ROC CURVES FOR INDIVIDUAL SUBJECTS .............................................................................. 61
FIGURE 22 - MEAN PERFORMANCE OF PARTICIPANTS (AREA UNDER ROC CURVES) AT DIFFERENT HORIZONTAL DISPARITY ................................................................................................................................. 64
FIGURE 23 - EXPECTED (HYPOTHETICAL) RESULTS FOR EXPERIMENT 2 .................................................. 67
FIGURE 24 - EXPERIMENTAL STIMULUS FOR EXPERIMENT 2 ........................................................................ 68
FIGURE 25 - TOLERANCE OF VERTICAL DISPARITY .................................................................................... 72
List of Appendices

APPENDIX A - EXPERIMENT 1 - SUPPLEMENTARY MATERIAL ......................................................... 82
APPENDIX B - EXPERIMENT 1: ANOVA .................................................................................. 85
APPENDIX C - EXPERIMENT 1: ROC CURVES ........................................................................ 86
APPENDIX D - EXPERIMENT 2: INTERPOLATION GRAPHS ..................................................... 93
1 Introduction

Aerial search and rescue (SAR) is an extremely important responsibility of the Canadian military. Each year, Canadian SAR units co-ordinate and conduct aerial operations over a huge expanse, comprising different land and marine environments. These operations include, for example, searches conducted over the northern part of Canada for downed aircraft. For essentially all operations, time is crucial to the survival of those involved in the accident. By searching from the air, the rescuers can cover a large area of rugged and often inaccessible terrain in a relatively short period of time.

In current operations, SAR technicians typically conduct aerial searches visually with their naked eyes, using binoculars as an aid when needed. One of the key features which help technicians detect a possible crash site is the broken treetops in the foliage caused during the crash. The main objective of this thesis was to study the possibility of using stereoscopic video as a means to assist SAR technicians in detecting such targets. In particular, the concept of hyperstereopsis was investigated as potentially useful for SAR operations. The main characteristic of a hyperstereoscopic video system is an exaggerated camera separation, which causes a corresponding increase in horizontal disparity in the stereo pair presented to the viewer. This results in the potential benefits of an enhanced depth resolution in the video system. In short, it was anticipated that hyperstereoscopic video can serve to exaggerate the relative depth separation of individual objects in the visual scene and thus enhance operators' ability to discriminate between them during SAR operations.

The objective of this thesis can be divided into two major components. These are:

1. **Practical application:** To design and construct a prototype of a variable baseline stereo video system capable of providing large separation stereoscopic images according to the crucial constraints in the Canadian aerial SAR environment.

2. **Theoretical studies:** Using the new stereo video system, to study the theoretical advantages of hyperstereopsis in a visual detection task.
Figure 1 provides an illustration of how the different sub-components of this thesis relate to one another. As shown, establishing the advantages of enhanced depth perception provided by hyperstereopsis is one of the underlying motivations for this thesis. The ensuing development of the stereoscopic video system was based on the goal of providing a flexible platform for exploiting these benefits in the aerial SAR environment, where it is believed that the consequent enhanced depth perception should be able to help technicians in the accurate identification of camouflaged crash sites within the foliage. Using the resulting prototype, this theoretical advantage was tested with respect to two key variables which were believed to affect depth perception using the new system. Two laboratory experiments were conducted. The first of these tested the basic theoretical benefits of hyperstereopsis, where it was hypothesised that detection performance (for simple stimuli) would increase as the magnitude of horizontal disparity (HD) increases, up to the point at which stereoscopic perception breaks down. The key objective of the second experiment was to study tolerance to vertical disparity, in light of the fact that vertical disparity (VD) is well known to be disruptive for stereoscopic viewing. With increased horizontal disparity, furthermore, it was hypothesised that tolerance for VD will decrease as a result of less stable binocular fusion.
Figure 1 - Thesis Overview

Figure 1 was created to illustrate how the different components of this report are related, but it also serves the secondary function of outlining the structure of the thesis. The theoretical background section which follows provides a more detailed discussion of the theoretical motivation for this thesis. It is followed by a complete description of the design and prototyping process for the new stereoscopic video system. That section is then followed by a detailed discussion of the laboratory experiments conducted to test the original hypotheses. Finally, to complete the cycle, the conclusions and suggestions for other studies are presented.
2 Theoretical Background

The use of stereoscopic video displays in aerial SAR environments is a concept that has not been thoroughly explored, based on the fact that a search through the literature yielded little material related directly to the use of stereo video in SAR. Nevertheless, the concept is based on sound perceptual theories and has the potential to be beneficial to performance not only in aerial SAR but also aerial reconnaissance. These theories have been applied to various other types of visual detection tasks and other practical teleoperation applications and have lead to improved performance. In other remote sensing applications, stereo viewing provides depth perception which often allows operators to perform tasks that would in fact otherwise be impossible. The understanding of these theories should thus provide many important design guidelines, as well as evaluation criteria, for the complete system.

2.1 Common Stereoscopic Display Applications

Currently, stereoscopic displays are commonly found in applications such as telerobotics and virtual environment systems. In virtual environments, stereoscopic displays are used to present 3D computer graphics, primarily as a means of increasing the sense of presence while the user interacts with the computer generated environment. Many studies have been conducted to explore different issues related to realism and sense of presence in virtual environments [Welch et al. 1996; Hendrix & Barfield 1996]. More specifically, the study conducted by Hendrix and Barfield on visual parameters of displays found that stereopsis has a positive contribution for the sense of presence in virtual environments.

Another popular area of application for stereoscopic display is in teleoperation, where, due to reasons such as hazardous environmental conditions, it is often necessary to complete robotic tasks remotely. In contrast to the computer generated images produced for virtual environments, stereoscopic displays are used in teleoperation to display live video images captured at the remote site. This type of system is thus more similar to that required for the present project. Many studies have been carried out to show that stereopsis is a beneficial addition to such visual displays [e.g. Tachi & Yasuda 1993; Drascic 1991; Cole et. al. 1991]. The findings in those studies have been consistent: with stereoscopic displays operators are able to complete their tasks faster and with higher accuracy.
Virtual environments and teleoperation are just two of the many different areas where the application of stereoscopic displays has proven to be beneficial. In those applications, stereoscopic presentation provides both added realism for visualisation and improved depth perception for manipulation. For the present project, we have explored yet another aspect of stereoscopic displays, for improving human performance in visual detection tasks. More specifically, the project's objective is to demonstrate the potential benefits of using stereoscopic video for detection of crash sites in a SAR operating environment, using a new type of single camera hyper-stereoscopic display system.

2.2 Aerial Photogrammetry

A search of the literature suggested that the use of stereoscopic video technology in aerial SAR and reconnaissance is virtually non-existent, even though stereoscopic displays as a general tool to present information has been quite widely accepted and used in similar applications for quite some time. That is, for many years the field of aerial photogrammetry has used stereoscopic photographs to provide useful information to surveyors and urban planners, such as in the production of topographic maps. Early prototypes in this area [e.g. Panton 1978; Young & Isbell 1978] included the use of digital stereo images in automated production of topographic maps. Using aerial images taken from two different locations, those systems extracted information about the environment that was unavailable from a single viewpoint. Unlike the stereo video system proposed here, however, photogrammetry systems typically use still images, instead of live video. In either case, in order to generate a three-dimensional image, at least two different images of the survey area must be available. Still pictures are usually taken from a camera pointing vertically downwards towards the survey area. As the aircraft travels, successive photographs are taken and used to create the stereo pairs. The single camera stereoscopic video system described in this report is similar to that photogrammetry technique.

Topographic map making is only one of the uses of stereo aerial photogrammetry. Other applications include geological and forestry surveys and urban planning. For example, stereo photographs have been used for the purpose of recording urban development of a particular area over time. Stereoscopic photographs are also used to make accurate height measurement of different structures within a large area, provided that certain parameters are carefully controlled while taking the required photographs. These parameters include the altitude of the aircraft and
the separation at which the stereo photographs are taken. The effects of manipulating these parameters on the precision of height estimates can manifest themselves in two ways. Firstly, the actual setting of these parameters can influence the effective "depth resolution" of the resultant stereoscopic photographs, an effect that can be calculated analytically based on the lens optics and basic geometry. Studies on these parameters have indicated that using a lower flight altitude and wider separation between the locations of pairs of pictures generally improves the depth resolution of the photographs, resulting in more precise estimates of object heights in the photos (Avery 1985). The second way these parameters affect precision is through their effects on the human viewer of the photographs, in terms of perceived quality of the photographs and viewing comfort.

Analogous to the photogrammetry studies outlined above, flight altitude and effective camera separation are also two of the key parameters affecting the performance of the proposed video system. The theoretical background to the perceptual issues related to these parameters are further explored later in this section.

2.3 Depth Perception in SAR Operations

During an aerial operation, one of the SAR technician's principal functions is to search the visual scene for a number of different cues that may indicate the possible presence of the target being sought. Depending on the mission, the available cues can be very diverse. When searching over a large body of water for stranded vessels, for example, targets are generally quite distinct from the surroundings, where there are few or no visual obstruction under favourable weather conditions. A similar example of SAR is searching over a body of water for survivors of a downed aircraft or sunken vessel. In such situations, the key visual cue may be the colour differences between the target and the background (e.g. bright orange personal flotation devices or a lifeboat against a blue background). In both of these cases, two major factors are involved in successfully performing the search task: 1) human visual acuity and 2) effective coverage of the search area. To enhance visual acuity, binoculars are used regularly during SAR operations. To deal with the second factor, a study funded by DCEM (Stager 1974) was conducted to design visual scanning patterns to provide guidelines for effective coverage of a search area.
The single camera stereo video system developed here was not designed for use in such high contrast situations. The system is intended, rather, to be an aid for SAR operations performed over areas of dense foliage, where search targets, such as the crash site of a downed aircraft, are often occluded and where many of the more obvious visual cues, such as fire and smoke, tend to dissipate relatively quickly after the incident. In order to understand the task domain of the SAR technician, it was necessary to interview technicians for their inputs. In January 1997, a visit was made to Canadian Forces Base (CFB) Trenton (SAR operation headquarters) to speak with the SAR technicians. During the interview, along with some information that is crucial to the design of the system (which is presented in a later section of this thesis), SAR technicians identified the presence of breakage in the tree line as one of the most important visual cues frequently used for detection of such targets. Unfortunately such broken treetops contain very little information in the form of contrast gradients or other easily detectable cues. Instead, one of the key cues attended to are changes in the topology of the forest cover, that is, changes in its three dimensional contour. The system presented here was consequently designed to help SAR technicians in detecting this type of target.

With current search procedures, technicians typically perform their visual searches either unaided or with the help of binoculars, for both detection and verification of possible targets. Due to the altitudes from which searches are conducted, however, targets are typically too far away for binocular disparity to be an effective cue. SAR technicians thus have to rely instead on such depth cues as interposition and shadows for detection, to make up for the effective loss of stereopsis. The present hyperstereoscopic video system addresses the absence of binocular disparity by providing the SAR technicians with this compelling depth cue through "artificial" means.

To understand some of the technical considerations for the design of the video system, it is helpful to review some of the theories associated with depth perception, especially as they pertain to the context of aerial search and rescue. The perception of depth is based on the interpretation of a series of cues present both within the environment and within the visual system of the observer. Interposition and motion parallax are two of the dominant environmental cues. Other, more subtle, depth cues within the scene include relative sizes of objects, texture gradient, and aerial perspective, among others. Feedback within the observer's visual system can also play a crucial role in the perception of depth, providing what is sometimes referred to as observer-
centred cues [Wickens 1992]. The three observer-centred depth cues are: binocular disparity, binocular convergence, and ocular accommodation. All of these cues are explained in more detail below.

### 2.3.1 Environmental Depth Cues

Reviewing first the non observer-centred cues, with current procedures according to the description above, the major environmental cues available to the SAR technicians are interposition, motion parallax and shadows. In a 3D scene, *interposition*, or *occlusion*, refers to the fact that an object closer to the observer can obstruct the view of another object that is further away, creating one of the most compelling depth cues. However, in situations in which the boundaries of different objects are not distinct, as in aerial search over forests, the effects of this depth cue are greatly reduced.

*Motion parallax* can also be a compelling cue for interpretation of relative position in a scene. In aerial SAR, as the aircraft travels through a search area, the line of sight of the technician looking downwards is generally perpendicular to the motion of the aircraft. The relative motion between the observer and the environment thus causes the scene to stream by approximately horizontally. Consequently, as the aircraft travels between two locations in space, an object in the scene that is closer to the observer relative to the transverse axis of the aircraft will appear to have travelled a larger distance per unit time than one that is further away. This apparent relative motion between objects on different depth planes, also referred to as motion parallax, can also be one of the most compelling cues in the environment. However, apparent relative motion is highly dependent on the ratio between the relative distance between objects and the viewing distance of the observer. If the observer is examining two objects at similar depths from a great distance, as in the case of aerial SAR, the motion parallax between the two objects will be very small.

*Shadows* can also be a strong cue for detecting breaks in the tree line. The drawback of this cue, however, is its variability under different lighting conditions. This cue is also extremely susceptible to other changes in viewing conditions, such as the presence of fog and overcast conditions.
2.3.2 Convergence and Accommodation Cues

Other than using cues in the environment, the human visual system possesses a number of other mechanisms which provide information to enhance depth perception. Convergence and accommodation of the eyes and binocular disparity all provide information which can then be interpreted by the human brain to create a three dimensional image of the surroundings.

Convergence refers to the situation in which the observer fixates on a particular object or area in the environment. In looking at the object, both eyes rotate inwards to place the object of interest on each fovea, which is the region of the retina that has the highest concentration of cone receptors and thus the highest resolving power. At the same time, the movement of the muscles surrounding the eyes gives feedback to the brain. Similar to receiving proprioceptive feedback on movement of limbs, the actual convergence movements of the muscles controlling the eyes are perceived by the brain and incorporated into perception of distance of the object from the viewer.

Accommodation is the related focusing of the lenses to view an object in the environment. In order to project a clear image onto the retina, the shape of the lens is altered by contracting the appropriate muscles in the eye. These subtle changes in the lens act as feedback to the brain and provide information regarding the position of attended objects in the surrounding environment.

2.3.3 Binocular Disparity

Binocular disparity is another important observer centred depth cue that is used to resolve ambiguity related to the spatial arrangement of objects in a scene. Having two eyes at two different horizontal locations provides two different retinal projections of the surrounding environment, from two slightly different viewpoints. The differences of retinal images from these viewpoints are interpreted by the visual system and fused into a single coherent percept of three dimensional space.

---

1 Cones are specialised receptors on the retina which are highly sensitive to colour and can resolve visual scenes in great detail. The other type of receptors found in the retina, the rods, are highly active in low light environment and reactive to motion in the visual field.
Figure 2 - Illustration of binocular disparity. An object, in this case a white circle, at a depth plane different than the point of convergence, will produce a binocular disparity in the left and right retinal projections. The left figure, the eyes are converged on the black dot inside a square surround. The right figure shows the retinal images of both objects on each eye. The magnitude of disparity = $D$. 
Figure 2 shows how two objects at different depth planes will be registered on the retinas of the observer. In this example, the square with the black dot in the centre is at the point of convergence. The image of an object on this depth plane will be projected on the same corresponding region of both foveae creating zero binocular disparity. In this illustration, the white circle is in front of the square, and thus in front of the convergence point. As a result of its location in depth, the image of the circle is projected onto a region right of the fovea of the left retina. At the same time, the image of the circle on the right retina is to the left of its foveal region. The horizontal disparity between the left and right retinal projections of the same object gives rise to the sensation of relative depth. In other words, because of the difference in retinal projections created by the two objects (the circle and the square), the viewer will conclude that the circle is in front of the cube.

Several studies in the literature have reported attributes of binocular disparity that may be of interest for the kind of stereo video display proposed here to be used in SAR operations. In Merritt (1991), the author identified visual noise filtering as one of the potential benefit of stereo displays. Under the assumption that the noise present within a scene is uncorrelated with the objects being examined, having two different views of the same scene may allow the brain to filter out the uncorrelated noise while more easily identifying the coherent target signal. In a report by Smith & Pepper (1981) of a comparative study of monoscopic versus stereoscopic displays for degraded viewing conditions, the authors concluded that stereoscopic video was less affected than monoscopic video by visual noise in the environment and under low contrast conditions. It is important to note that this type of condition is similar to the overcast or foggy conditions often encountered during SAR operations.

Another series of studies conducted by Schneider and colleagues (Schneider, Moraglia et al. 1989; Schneider and Moraglia 1994) proposed similar benefits of stereoscopic viewing. In
their studies, Schneider et al. identified an effect which they called *binocular unmasking*, whereby the use of two different views enhances a viewer's ability to detect a target camouflaged within the visual scene. Schneider et al. postulated that binocular summation thus plays a key role in the camouflage breaking ability of stereopsis.

In a related study, Ma (1999) investigated the benefits of stereoscopic video as a means of abetting camouflage detection underwater, under the particular constraints of turbidity. He showed not only that detection performance with stereoscopic video is generally greater than with monoscopic video, but also that the rate of decline of detection performance as turbidity levels increase is significantly less with stereoscopic than with monoscopic video [Ma 1999].

From the point of view of the present project, this camouflage breaking effect can clearly be of value to SAR technicians searching for targets camouflaged within dense foliage. Other research related to the practical advantages of binocular disparity include studies done on remote driving and hazard detection. In many of those studies, the experimental task involved visual detection of targets or hazards on a video display. In Merritt & CuQlock-Knopp (1991), participants of the experiment were asked to identify hazards within a video scene shown in either 2D or 3D. The results of that experiment also showed dramatic benefits with stereo video displays.

### 2.3.4 Hyperstereopsis

Building on the idea of binocular disparity, one of the design criteria for the stereoscopic video system developed here was the use of "hyperstereopsis" to further enhance visual search. Under normal direct viewing conditions in humans, the horizontal separation, or interocular distance (approx. 63.5 mm), between the left and right eyes causes two slightly different retinal projections on the eyes. (See Figure 2.) In a stereoscopic video system, the left and right cameras can be positioned to provide a comparable viewing condition, when combined with the appropriate optical elements and when the monitor is viewed from an appropriate distance. Whenever a stereoscopic display reproduces the conditions corresponding to normal viewing, this case is referred to as *orthostereoscopic* viewing. One of the reasons to use orthostereoscopy for teleoperation is that it provides the most natural view to the operator, allowing him to make accurate absolute depth judgements because the information provided is similar to that of direct viewing.
Whenever the equivalent camera separation is increased well beyond the normal interocular distance, however, the relative difference in binocular disparity between any two objects in the scene will also increase. This can serve to exaggerate the perceived relative depth differences between different objects, producing what is known as hyperstereopsis. Taking for example a scenario where a target to be detected is a hole on a surface, by increasing the separation of the two viewpoints, the hole will appear deeper than it would normally appear, thus creating a stronger signal for detection. For more information on differences between ortho-, hyper- and hypo-stereoscopy, the reader is advised to consult Diner & Fender (1993).

Since perceived depth in a hyperstereoscopic view is exaggerated, it is difficult to make absolute judgements, regarding actual locations of objects in depth or absolute dimensions or distances between objects, because their relationships with other landmarks, which would ordinarily be used for comparison, are distorted. However, for the purpose of detecting the presence of targets in depth, such exaggeration can be beneficial. In a study conducted by Spain (1989), the author presented the following two conclusions:

1. Both image magnification and increases in camera interaxial separation are useful factors for enhancing target detection time and recognition rate with stereoscopic TV systems.

2. The interactive effects of image magnification and variable camera interaxial separation are not disruptive of visual performance for remote reconnaissance tasks.

Other studies conducted on the effects of hyperstereopsis have shown mixed results. It would appear thus that the effect is highly dependent on the task to be performed, even though geometrically increases in separation do increase the depth resolution of a stereoscopic video system (within the limits of the pixel resolution of the CCD camera). The actual calculations for this effect can be found in Diner & Fender (1993).

From the definition given in Diner & Fender, any camera separation larger than normal interocular separation can be considered as equivalent to hyperstereoscopic viewing. Such a “definition” holds little practical meaning, however, especially in relation to SAR operations, because for actual search altitudes, which are typically in the range of 150 to 500 metres, a 63.5 mm separation corresponding to normal human eyes would yield little effect. The actual height
during a search operation is usually determined by the environment in which the search is to be carried out. Consequently, in order for a hyperstereoscopic viewing system to be effective for SAR, effective camera separation would obviously have to be very much greater than the minimum 63.5 mm separation defined by Diner & Fender. In conclusion, a useful stereo video assisted SAR system should provide the capability to scale to large search altitudes and different focal lengths while providing hyperstereoscopic viewing, in order to facilitate the important detection portion of the SAR technician's job.

2.3.5 Accommodation-Vergence (AV) Mismatch

In applications of stereo video displays in which a CRT is used, the accommodation and vergence cues discussed in Section 2.3.2 often provide conflicting information. In the most common type of stereo display implementation, the CRT display alternately presents left and right eye views to the corresponding eyes, using a pair of liquid crystal (LC) shutter glasses at the update rate of the CRT display. By synchronising the shutter glasses with the refresh rate of the CRT screen, it is possible to present the correct image to the appropriate eyes to obtain a sensation of depth. This is the type of display that was used for the hyperstereoscopic video system presented here.

The main perceptual problem with using a CRT display, that can not easily be corrected or minimised, is the accommodation-vergence (AV) mismatch. Whenever the left and right images, commonly referred to as a stereo pair, are presented, the information provided by convergence and that provided by accommodation are often not congruent, in the sense that the two sources of information do not match the expectations of the viewer created by past experience. Depending on the location of objects in the scene, the disparity created between the left and right images will cause the eyes typically to converge on a particular location, indicating that the object is either in front of or behind the monitor. However, since the actual picture is always presented on the screen itself, the lenses in the viewer's eyes must necessarily accommodate at the distance from the eyes and the screen to focus on its surface. The feedback mechanism associated with accommodation thus indicates that the object is located at the screen distance, thus conflicting with the vergence cue information.
The actual effect of the AV mismatch has been the topic of a number of different studies. In one of these, conducted by Gooding et al (1991), the results showed that perceived depth (with respect to the screen) is generally lower than that predicted by geometry. A possible explanation of this finding is that, even though binocular disparity and convergence (both created by the geometry of the displayed image) together place the object at particular distance away from the plane of the screen, the ocular accommodation at the screen distance acts to reduce the overall perceived distance. This effect may be related to the weighted additive model of depth perception described in a report by Wickens, Todd and Seidler (1989). According to that model, the effects of different depth cues and their relative importance in depth perception are described. With respect to AV mismatch, it can be postulated that this mismatch is a negative factor in the additive sum. In Gooding et al, the experiment showed that, as the viewing distance increased, the ratio of actual perceived depth to that predicted geometrically became closer to one. This can be expected since the AV mismatch is reduced as viewing distance increases.

2.4 SAR Visual Detection

In most of the published visual detection studies conducted for aerial environments, the main interest has been detection performance of pilots in tactical and or reconnaissance environments, rather than search and rescue carried out by non-pilot observers. The study conducted by Stager for DCiem, which concentrated on SAR type detection, is thus an exception (Stager 1978). In his simulation of a search task, which incorporated the procedures being employed at the time by the Canadian SAR unit, Stager found that the actual visual coverage of SAR technicians is quite low. Stager concluded that, under the time constraints of the task, it would be essentially impossible for a single observer to provide adequate coverage of the search site, and he thus recommended having two searchers on either side of the aircraft, to improve effective area coverage.

Although Stager's recommendation was adopted by the Canadian military and has improved detection performance, the present project commenced with the premise that, by using a stereoscopic video system, it should be possible to increase the effective number of observers performing a search even further, this would only be limited by the number of well trained SAR
technicians available for a SAR operation. Furthermore, following the suggestions of Spain (1989), it was postulated that detection performance might be improved even more if hyperstereoscopic viewing were to be provided to those observers. Finally, it was hypothesised that, if hyperstereopsis can help unmask visual targets, it may also help increase speed of target recognition.
3 System Design & Prototype

As part of the project, a stereoscopic video system was designed and a prototype built specifically for aerial SAR applications. The final design chosen was a single camera hyperstereoscopic video system, which takes advantage of the motion of the aircraft to create stereo pairs using only a single camera. The details of the chosen design and alternatives selected are described in the present section.

In order to better define the design, two separate trips were taken to gather data pertinent to the project. The first trip was to Trenton, Ontario, where the home base of SAR operations covering central Canada is located. The main purpose of that trip was to interview a SAR crew to find out about the responsibilities of the crew and other important information regarding actual search flights. (See Section 3.2.) This information was used for the design of the software and hardware of the stereo video system. After the completion of a running prototype of the software, a second trip was taken to Kingston, Ontario, to gather video data for testing purposes. (See Section 3.5.)

3.1 Design Considerations

At the onset of the project, two potential designs were considered for the aerial Search and Rescue stereoscopic video application: 1) a dual camera stereo system and 2) a single camera motion stereo system. The first alternative is similar to typical stereoscopic video systems used in teleoperation and other applications. The second alternative takes advantage of the motion of the search vehicle to generate the binocular disparity necessary for stereoscopic viewing. For both systems it was decided that binocular viewing would be achieved with the aid of liquid crystal (LC) shutter glasses, where the left and right eye images are presented on a CRT display and the left and right eye LC shutters are synchronised to the refresh rate of the display.

3.1.1 Dual Camera Stereo Alternative

In a dual camera stereo system, the two cameras are mounted pointing downwards with a horizontal separation (with respect to the orientation of the CCDs) on a level platform. The
distance between the two cameras is referred to as the camera separation. The convergence angle refers to the angle at which the cameras are mounted relative to each other and can be altered according to the application. This allows for a converging geometry that mimics the convergence of the human visual system. These two parameters are critical for predicting the depth resolution of the system. Furthermore, performance with such systems as well as subjective acceptance of them are both highly dependent on these factors.

The major advantage of such systems is that the simultaneous availability of the two images of the stereo pair allows multiplexing of the signal using simple hardware circuits. This ensures that the video image can be updated in real time during flight. This type of system also allows easy converging of the cameras, as illustrated in Figure 3. This may offer some perceptual advantages because of its similarity to the human vision system. The actual stereo region of the video image can also be maximised because of the converging cameras, as explained in Section 3.4.1.

There are also disadvantages associated with using this type of system for SAR applications, however. First of all, the system requires two cameras and, therefore, two camera mounts. In order to minimise the effect of mechanical vibrations on video quality and vertical disparities, the camera mounts must be designed specifically for this type of application. Gyroscopic camera mounts for aerial video are currently available commercially; however, the cost of such mounts is very high and two such camera mounts would make the system extremely expensive. Furthermore, these camera mounts must be attached to the exterior of the aircraft. Not only would two such secure sites be required but, to add to the problem, the additional wiring required would also make this option expensive and thus less desirable.
Another important drawback of dual camera systems is the limitation they impose on the separation of the cameras. To achieve the required hyperstereoscopic viewing, the cameras must be mounted a relatively large distance apart. Given that searches are to be conducted from an aircraft, one option would be to mount cameras near the wing tips, with their horizontal axes parallel to the wing axes, thus achieving a separation approximately equal to the wingspan of the aircraft. Although this might create a very large camera separation, the flexibility of the wing is likely to be a severe problem. For example, the wing tips of the Hercules aircraft can flex up to 2 meters during flight. Although on the one hand a change of 2 meters in linear displacement of the cameras is unlikely to have a large effect on image quality, given the search altitude, the fluctuations in the relative angle at which the cameras are pointed is a much more critical problem. Since flexion of the wings creates a rotating motion around the fuselage at the fulcrum, such motion can cause the optical axes of the cameras to deviate greatly from their desired orientation. Furthermore, wiring is also a major problem with mounting cameras on the wings, since having to pass wires within the wings makes this option impractical.

Figure 3 - Dual Camera Stereo System with Converging Geometry.
The other dual camera option is to mount the cameras along the centre line of the aircraft at opposite ends of the fuselage, with their horizontal axes parallel to the longitudinal axis of the aircraft. Whereas this would help eliminate the problem created by the flexing of the wings, this option still suffers from the same problems of limited camera separation, since the maximum distance between the two cameras will approximately equal the length of the aircraft.

Another technical problem with conventional dual camera systems for this particular application is adjustment of camera separation. In a typical video application, the two cameras are mounted on a single camera platform. In most situations adjustments to camera separation are relatively infrequent, but when they are necessary, such as for adjusting depth resolution or fusible field of view, adjustments can be made by manipulating the camera mount directly. This presents a problem with cameras mounted outside of the aircraft, however, since, to change the camera separation, the physical camera mounts would have to be able to move along the centre line of the fuselage.

The alternative is to select an acceptable fixed separation, which may however be sub-optimal. That is, whereas it may be possible to select a separation that will produce a sufficient disparity at typical search altitudes, at the same time, the separation must allow operator to fuse all images without eyestrain. With individual differences between operators, if this approach is taken, the separation used will likely be conservative (limited by the physical characteristics of the aircraft), which could reduce any performance advantages of the hyperstereoscopic video system. In such cases, the convergence angle of the cameras can be adjusted to increase the stereoscopic portion of the captured image and depth resolution can also be altered to some extent; however, the range of resolution will be limited.

A final consideration for dual camera systems is the effect of visual flow on the CRT display. Under most circumstances, optical flow patterns can be classified as comprising some combination of radial expansion and lateral optical flow, as illustrated in Figure 4. Due to the inevitable presence of crosswinds, the aircraft's actual path, that is, its ground speed vector, will rarely be aligned with the baseline of the two cameras. Taking the case of longitudinally mounted cameras for example, as illustrated in Figure 5, whenever horizontal ground speed motions are present due to crosswinds, objects which should appear to be moving horizontally across the video screen will acquire a vertical component. This will cause not only individually
discernible objects but also textures, such as foliage, to appear to travel diagonally across the screen. Operators may thus have difficulty adjusting to such high frequency textures flowing across the display in an unfamiliar direction. For example, if the optical flow were to acquire a large vertical component, instead of horizontal as expected, the resulting optical texture flow could cause disorientation.

Figure 4 - Typical optical flow pattern encountered in daily life.

Figure 5 - Diagonal movement of objects on ground across the screen in the presence of crosswind. Note that cameras are mounted with their horizontal axes parallel to longitudinal axis of aircraft.
3.1.2 Single Camera Motion Stereo Alternative

As an alternative to the traditional dual camera stereoscopic video system, a motion based single camera stereo system was considered for the project. In order to generate stereoscopic video, two images of the same scene must be taken at two different physical locations to create the necessary binocular disparity. In some single camera systems, this is feasible through the use of a series of mirrors which reflect the images from two different locations to the CCD of the camera. The large advantage of the SAR scenario is that the complication of mirrors and lenses can be eliminated because the search is conducted from a moving vehicle. To generate the two images needed for a stereo pair, the camera can capture one image and store it. As time elapses, the vehicle and the camera move to a new location at which point another image can be captured. Assuming that the search vehicle does not alter its course drastically over this short period of time, the two images captured at the two different locations can be used to create the necessary stereo pair.

One of the main advantages of a single camera system is the simplicity of the mechanical hardware required to implement the system, since only one camera mount is required. This also simplifies the requirement of wiring and accessibility to the exterior of the aircraft.

Another major advantage of a single camera stereo system is the flexibility of the camera separation. Unlike a dual camera system, the cameras capturing each of the two images needed for the stereo pair are not separated in space. Rather, the same camera is used to capture both images by moving the camera from one location to another. In a sense, therefore, the "cameras" are in fact separated, but in time rather than in space. The physical separation between the locations at which the two images are taken is referred to as the effective camera separation. Although theoretically the effective camera separation has an upper limit, imposed by the field of view of the resulting fused image, and a lower limit, imposed by the distance needed to be travelled by the vehicle during the time between successive frames, with this system the effective separation is no longer limited by the physical dimensions of the vehicle, as is the case for example in Figure 5. This has the consequence that a much larger variety of different aircraft can
be used for search and rescue, including helicopters. It also provides the advantage of being able to change the effective camera configuration in software as necessary, rather than via hardware, as explained below.

In order to create a stereo pair using two images "separated in time," a large frame buffer is necessary to store the images. The primary choice for implementing this buffer is to use the main memory of a computer workstation. The monitor can then be used to display the video image and the computer can be programmed to produce the necessary control interface for the operator. Since the actual time delay needed is created using software, it is relatively easy to change the effective camera separation simply by changing the time delay between the two images used for the stereo pair. The system can also allow for proper re-scaling of effective camera separation dynamically, as the search altitude or speed for the mission changes. Conversely, the same camera configuration designed for a particular hyperstereoscopic set-up should be readily maintainable even if the altitude or ground speed changes, in response to monitoring of those parameters. Finally, with such a system individual operators should be able to adjust effective separation to reduce any eyestrain while maintaining binocular fusion, even if several operators are sharing the same single video camera image (but are using different display computers). Since the actual viewpoint separation is controlled by the software, it is possible for the same video signal to be shared by multiple computer systems each having a different horizontal disparity level.

There are also certain shortcomings with a single camera system, however. Since the stereo video is generated using a time-delay and the motion of the aircraft, the stereo image parameters can not be controlled strictly in hardware. Because it must be done in software, this can present a problem since the computer system may not have the power necessary to process the dynamic time delays and create delayed video at full frame rate.

Another disadvantage of the system is the intrinsic lack of convergence in the effective stereo pair. Since only one camera is used, the simplest configuration is to aim the camera directly away from the aircraft, resulting in a situation similar to that of a dual camera stereo system with parallel optical axes. Although it is not typically a major problem for viewers to fuse left and right eye images in such cases, it does reduce the shared portion of the image that contains fused information in stereo.
In light of its advantages, which appear to outweigh its disadvantages, the single camera stereo concept, with parallel optical geometry, was chosen for implementation in this project. The technical implementation and some of the detailed considerations in its design are discussed in sections 3.3 and 3.4 of this report.

3.2 SAR Technician Interview: Trenton Field Trip

The main objectives for the first of two field trips taken for the project were to gather information on the current mode of SAR operations and to present some preliminary ideas to the SAR technicians. With the help of the scientific monitor of the DCIEM contract, a visit was arranged to the home base of SAR operations at Trenton, Ontario -- one of two such units operating in Canada. The visit began with a short presentation of the concept of using stereo video as an aid to SAR technicians in their tasks. A list of questions was prepared to gather information about the tasks and responsibilities of SAR technicians, the equipment that is currently used as aids, the search vehicles used and current operating conditions. The answers to these questions were crucial to the design of the video system. The important issues investigated are described in the following sections.

3.2.1 Search Vehicle

At the time of the Trenton visit, the Canadian Forces employed three different aircraft for its SAR operations: 1) the CH-13 Labrador helicopter, 2) the CC-115 Buffalo, and 3) the CC-130 Hercules transport. Some of the important aircraft related characteristics pertaining to SAR, such as range, speed and capacity, are summarised here.

**CH-113 Labrador Helicopter:**

- Range: 1,131 km
- Max Speed: 125 knots (cruising)
- Search Speed: ~80 knots
- Payload: 18 passengers, or up to 2,727 kg

Although the Labrador helicopter has since been retired from service, its main advantage of the Labrador for SAR operations is its ability to hover and conduct searches at low speeds and
altitudes. Its hull is water tight, allowing for water landings. It is also equipped with a rescue hoist for lowering SAR technicians and lifting survivors to safety.

**CC-115 Buffalo**

Range: 1,886 km  
Max. Speed: 220 knots (cruising)  
Payload: 41 passengers, or max. 2,727 kg

**CC-130 Hercules Transport:**

- Range: 3,960 - 9,790 km  
- Max Speed: 300 knots (cruising)  
- Search Speed: As low as 125 knots  
- Payload: 92 passengers, or 17,340 kg

Its long operating range and its ability to carry a large team of SAR technicians makes the Hercules the primary vehicle used for rescue operations by the Canadian Forces. The stereo video system described in this report is designed mainly to be used on the Hercules. Among its other advantages, this massive aircraft has the payload capacity and the AC power required to operate the different pieces of equipment included in the stereo video system.

### 3.2.2 Search Task

During a SAR mission, the main responsibility of the SAR technician is to perform visual search of the terrain below, to locate and identify the target sought. Two large bubble shaped windows located near the tail of the Hercules aircraft\(^2\) allow the SAR technicians to perform visual search looking outwards from the sides of the aircraft as it flies over the terrain. With the ability to carry a large SAR crew, the technicians can take relatively short shifts with substantial rest periods between shifts to maintain vigilance and avoid fatigue.

The technicians perform their visual searches either by naked eye or with the help of binoculars. The target of a visual search can comprise one or more of a number of different visual

\(^2\) Since the present system was designed primarily with the Hercules aircraft in mind, discussion of the search task is limited here to operations with this aircraft.
cues, including colours, smoke and broken treetops. During the interview with the technicians, it was learned that the search pattern used is not a continuous sweep of the target region. Instead, it comprises a series of fixations illustrated by the squares shown in Figure 6 relative to the aircraft frame of reference\(^3\). Each individual square in Figure 6 represents a single fixation in the search area. The dashed-line arrows indicate the sequence in which these fixations are supposed to take place. The squares are also numbered to illustrate how these fixations, measured in the frame of reference of the SAR technician, translate to the actual coverage of the ground. When translated into fixations relative to the ground frame of reference as the aircraft flies by, the visual field movement (resulting from movement of the eyes and/or the head) produces a pattern of visual sweeps perpendicular to the axis of the aircraft, as shown in Figure 7.

If the SAR technician spots a possible target, he/she signals the flight crew and the aircraft then can circle around for confirmation. During the interview, the technicians commented that, even though circling around can be time consuming and could potentially delay the mission (which can often last for several hours on the Hercules), it is absolutely crucial to verify any doubts that they may have before continuing with the mission.

![Figure 6 - Recommended visual search pattern, in aircraft frame of reference, based on only head and/or eye fixation of SAR tech. Numbers refer to successive fixations.](image)

\(^3\) This visual search pattern was recommended by Dr. Paul Stager in 1974 (Stager 1974), as a result of a study conducted for DCIEM.)
Figure 7 - Resulting visual sweep of target area from Fig. 6, in ground co-ordinates. (Numbers correspond to fixations shown in Fig. 6, in aircraft frame of reference.)

Depending on visibility and other conditions, typical flight altitudes vary from just over 500 ft to 1,500 ft. Whenever visibility is above 3 miles, the search is generally conducted at 1,500 ft. If visibility deteriorates, however, or if the weather ceiling is low, the crew may operate at about 500 ft. To cover a particular target area, the aircraft will make a series of parallel passes over the region, as shown in Figure 8. The separation between the parallel passes depends on the altitude flown, and thus on the visibility.
In the preceding discussion, some of the initial advantages and disadvantages of both the single and dual camera systems were identified. The main properties of the two designs are summarised and compared here in Table 1. The chief objective of the Trenton visit was to gather information regarding the current state of actual SAR operations. Using this information, a specific design was then selected for more detailed consideration and implementation.
<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dual Camera Stereo</td>
<td>Requires 2 camera mounts</td>
</tr>
<tr>
<td>• Stereo video at full frame rate achieved with hardware</td>
<td>• Extensive wiring and complex mechanical hardware necessary.</td>
</tr>
<tr>
<td>• Converging geometry allows full stereo field of view (see Section 3.4.1)</td>
<td>• Difficult to scale camera separation to altitude changes</td>
</tr>
<tr>
<td></td>
<td>• Limitation of maximum camera separation</td>
</tr>
<tr>
<td>Single Camera Stereo</td>
<td>Parallel geometry will limit stereo field of view</td>
</tr>
<tr>
<td>• Single camera, single mount</td>
<td>Implementing stereo video in software will reduce frame rate of video</td>
</tr>
<tr>
<td>• Software stereo allows simple and flexible scaling to altitude changes</td>
<td></td>
</tr>
<tr>
<td>• Effective camera separation not limited by dimensions of aircraft</td>
<td></td>
</tr>
<tr>
<td>• Minimal mechanical hardware</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 - Comparison between dual and single camera designs.

During the Trenton visit, interviews with the SAR crew helped identify some crucial criteria in selecting the proper stereo video system. These criteria included the tasks and equipment discussed above, as well as the availability of hardware and power on board the search vehicle. Another important set of criteria that did not arise directly from the Trenton discussions related to choice of stereoscopic video technology. Because of the theoretical advantages of
hyperstereopsis, as summarised in Table 1, the single camera system was selected as the design for implementation.

3.4 Selected Design

After consideration of the stated alternatives, the design selected for implementation was a single camera system which generate stereoscopic video using a delaying buffer. This decision to choose single camera implementation over a dual camera system was made based on several different factors. The simplest explanation is that the pros outnumbered the cons for the single camera system, which was not the case for the dual camera system. However, the decision was based on several other important factors as well. One of the major drawbacks which made dual camera undesirable was the hardware complexity involved in realising such a system. In order for implementation, two cameras with corresponding mounting points and additional wiring would be needed. Another key factor for the selection was the difficulty which would have been involved with changing camera separation with the two camera system, since it would have required at least one camera mount to be mobile. Keeping in mind that this would have had to be done on the exterior of the body, this would have been very difficult to realise. Furthermore, the reliability of such a mobile mount may be questionable due to expected rapid temperature changes and exposure to moisture.

Although a single camera system can avoid all of the above shortcomings, this is not to say that the single camera system is the perfect choice. As summarised in Table 1, the single camera system would reduce the stereo field of view and full frame rates (at least for the first prototype) were not expected. However, the flexibility, stereo scalability and hardware simplicity made it the more obvious choice. In the following sections, some of the crucial specifications are discussed in more detail.

3.4.1 Design Specifications

In this section, the requirements of the selected design are discussed. These requirements defined the specifications of both the hardware and software portion of this project.

- Using motion to generate stereo video:
Since the system is designed to generate stereo video through horizontal binocular disparity using the motion of the search vehicle, the camera must be mounted in such a way that the video image will flow horizontally across the viewing screen. Figure 9 illustrates the essence of this requirement. To achieve this, the camera must thus be mounted with its vertical axis perpendicular to and its horizontal axis parallel to the direction of motion of the aircraft.

Assuming that the aircraft flies along a straight line relative to the ground, the stereo image will then appear to have been captured by a pair of cameras with parallel optical axes. Ordinarily, human operators should have very little trouble fusing such non-converging images. However, as illustrated in Figure 9, such an effectively parallel camera system will ordinarily limit the size of the effective stereo field available to the operator, since this corresponds to the common field comprising the resulting fused image shown in the shaded area. This shaded area is defined by the dashed-line triangular regions below each aircraft symbol, and illustrates the field of view of the camera mounted directly below the plane.

The two images in the lower portion of Figure 9 represent images captured by the camera at the two different locations. Because the camera is pointing vertically downwards, the images captured at different points in time will only partially overlap. The resulting common fused stereo region, shown shaded in grey in the illustration is obviously smaller than each individual camera FOV. This is not the case with a traditional dual camera system, however, which can be set to point inwards towards a point of convergence and generate an almost 100 percent overlapping region between the two images.
Minimising Vertical Disparity:

Stereo video systems allow depth perception by means of horizontal binocular disparity. Whereas horizontal disparity is obviously beneficial, vertical disparity can often be a major problem in such systems. In a dual camera system, this problem can be minimised by carefully mounting the cameras on a level platform. In our single camera system, however, extra problems of vertical disparity are liable to be created due to the motion of the aircraft. During flight, disturbances such as cross winds will typically cause the aircraft to have some lateral component to its motion, and thus the actual heading of the aircraft relative to the ground is rarely the same as the direction of the fuselage. This will then translate into a vertical disparity between the delayed and current images, as illustrated in Figure 10. In general, users often report eyestrain and/or loss of binocular fusion with only a few pixels of vertical disparity of the screen.
depending on distance from the screen (~25 minutes of arc). This will also cause a reduction in stereoacuity as vertical disparity increases [Ogle 1955]. Depending on the flying conditions, the vertical disparity caused by cross winds could be several order of magnitude more severe than this, which would make it effectively impossible for the operator to fuse the left and right eye images unless appropriate measures are taken to prevent this phenomenon. This problem can be minimised by using an appropriate camera mount, as discussed in section 3.4.2.2 and in figure 11.

Figure 10 - Vertical Disparity resulting from lateral motion of aircraft
• **Controlling Effective Camera Separation**

Another important factor to consider in the single camera design is control of the effective camera separation. Since the system uses motion and a video buffer to create binocular disparity, any unaccommodated changes in ground speed for any reason will result in a change in the effective camera separation. Perceptually, this could be a problem for the operator, since it might affect perception of depth within the final image. Although there is some evidence in the literature that operators may not notice such changes (Milgram & Krüger, 1992), there has not been a lot of research carried out on this effect for this particular application and, in particular, on how it might affect detection performance in SAR. It is also conceivable that this could have a distracting effect on the operator, and may even elicit symptoms similar to those of motion sickness.

• **Re-scaling Effective Camera Separation to Changes in Search Altitude**

As identified in the Trenton visit, the actual search altitude of the SAR mission can change depending on the flight conditions and the visual details required. In order to achieve and maintain hyperstereopsis during such missions, the effective camera separation must be re-scaled as the altitude changes.

### 3.4.2 Hardware Design

The major hardware components for the system consist of a camera for video capturing, a mount which allows the camera to be attached to the exterior of the search vehicle, a Global Positioning System (GPS) receiver and a graphics workstation. Each of these components are described in more detail in the following paragraphs.

#### 3.4.2.1 Video Camera

The video capturing component is relatively simple for this single camera system, the basic requirement being essentially any simple video camera. The actual video signal can be one of the many established standards, such as NTSC or PAL. Depending on the camera mount used, as explained below, the physical requirements for the camera may change somewhat however. Specifically, if the camera mount does not have a vibration damping capability, then the camera
itself must be able to stabilise the video image. Currently, consumer CCD cameras from several major manufacturers, such as SONY and Panasonic, offer image stabilising capability by digital means. Although such cameras can digitally correct for high frequency movements created by an unsteady hand or while shooting video in a moving vehicle, it is not yet known whether such compensation would be adequate for aerial SAR operations.

3.4.2.2 Camera Mount

In order to have an unobstructed view of the terrain, it is important for the camera to be mounted on the exterior of the search vehicle. Although no actual camera mount was constructed as part of this project, several requirements for its design surfaced during the feasibility study and design phase.

- **Vibration Damping**

  One of the major problems with mounting a camera to a moving vehicle is the expected vibrations. This is particularly important for the single camera stereo system, where vibration in the system can cause serious vertical disparities between the left and right camera images. In the case of aerial SAR, for example, a small angular change in the mounting position can be magnified dramatically when the target can range from 500 to 1,500 feet away. As stated above, although it is possible to correct some of the effects of vibration to some extent digitally, with a good quality camera, it is much more desirable to do so with a properly designed, gyroscopically damped camera mount. During the Kingston test flight (Section 3.5), video images captured using such a system, the Wescam system (see www.wescam.com for additional information), were quite encouraging.

- **Rotation of Camera About its Optical Axis**

  In order to compensate for the lateral component of the aircraft's motion relative to the ground created by cross winds, the camera must be able to rotate about its optical axis. As shown in Figure 11, by rotating the camera, it should be possible to correct, at least partially, for the vertical disparity problem which results from lateral motion, as illustrated in Figure 10. In order to realise such a feature, however, it is also necessary to be able to track the ground speed vector accurately and be able to effect the rotational
changes with sufficient speed and accuracy. In fact, such a feature was available in the Wescam system used in the Kingston test flight (Section 3.5).

Figure 11 - Eliminating vertical disparity by rotation of camera about its optical axis. (Compare with figure 10)
• **Control of angle between optical axis of the camera and the ground**

Ideally, the camera mount should also provide users with the ability to change the camera pitch angle relative to the ground, as shown in Figure 12. Although this is not crucial for generating stereo video images with a single camera, such a panning feature could help increase the flexibility of the system. By aiming the angle closer to the horizon, the video flow across the screen would slow down, making it easier for operators to adapt. Such a correction is a standard feature in camera mounts like the Wescam unit used in the Kingston test flight and can easily be added to the single camera system.

![Image](image.png)

*Figure 12 - Sideways view vs. Straight down view.*

### 3.4.2.3 Computer workstation

A Silicon Graphics Inc. (SGI) Indigo 2 workstation was used to develop the software needed to implement the time delay video buffer. The SGI unit was chosen because of the large selection of C and C++ routines available. The workstation was equipped with one MIPS R4400 CPU, a Galileo video unit, which includes a Cosmo JPEG compression board, and 128 MB of memory (RAM). The Galileo Video unit allows both RGB and composite (e.g. NTSC and PAL) video signal as input sources. This unit also provides video output ports for composite output (NTSC or PAL) to different display devices, as well as in S-Video and RGB format.
3.4.2.4 Global Positioning System (GPS) Receiver

In order to control the effective camera separation dynamically, a GPS receiver unit was added to the system. With the GPS receiver, the distance between the locations where two different images are taken can be computed, either by comparing the GPS co-ordinates or by using the velocity information provided by the receiver. A 12-channel ALLSTAR GPS receiver, manufactured by Canadian Marconi, was selected for this project. The ALLSTAR receiver is an OEM board designed to be connected to a computer via a RS232 serial connection. For this project, the unit was set to operate as a stand alone receiver. With Selective Availability (SA) imposed, position accuracy of the unit in non-differential mode is less than 40 m (manufacturer's data) and velocity accuracy is < 0.2 m/s. Since the stereo video system requires only the relative location between two images to compute the effective camera separation, the GPS velocity information, given its superior accuracy, was deemed to be sufficient.

3.4.3 Software Design

By keeping the hardware relatively simple, the major development effort of the project focussed on creating the necessary software to create stereo video with a single camera. The software package was developed using a variety of C and C++ libraries specific to the SGI system. These libraries included VL (video library), Motif, X11 and others. The actual code of the software can be divided into three main modules (see figure 13).

The main goal of the first module was to capture the incoming video from the camera. To accomplish this, the hardware of the computer must be set to the correct configuration. With the particular Galileo video system on the Indigo 2, the hardware set-up takes on a path metaphor, meaning that the video signal travels between different "nodes" on a video path. A simple video path was established using a series of VL commands. The resulting path includes a single video source from the Galileo video board and a drain node in the main memory of the workstation. This approach allows the program to take video input from a video buffer and transfer the data into main memory for storage.

The same module is also responsible for setting up a time delay buffer for the single camera stereo system. A ring of frame buffers was used to implement the necessary time delay in the system. As video image frames become available from the source, the RGB data for each
pixel are copied into subsequent frames of the buffer. Once the last frame buffer is full, the first image stored in the ring is then overwritten, to allow continuous storage without overflowing the main memory into disk storage. Two pointers, L_image and R_image, are used to track the locations of the left- and right-eye images on the ring. The images stored at L_image and R_image locations are copied into the left and right video buffer respectively and then presented to the operator with the aid of the LC shutter glasses.

The second main section of the code is designed to handle the data from the GPS receiver, which is used by other portions of the code to calculate the required time delay for a chosen effective camera separation. The ALLSTAR GPS unit acquired for this project is capable of outputting NMEA messages via a standard RS232 serial port and can be connected to the workstation via one of its serial ports. The workstation can then extract the GPS information from the unit as required. The ALLSTAR GPS receiver can also provide different kinds of NMEA messages containing different information, where each single line message begins with a header containing information about the format of the remaining message.

The third and final section of the code is designed to provide an interface for the operator to interact with and control the single camera stereo video system. The interface is created for X windows using the X11 and Motif libraries for the SGI. The current interface is designed to be simple, with only the required basic functions to allow the operator to configure the system to his/her needs. Some of these functions include the ability to freeze a frame, for closer inspection of a particular area, and then return to live input. Another is the capability to increase or decrease the effective camera separation of the system manually, to allow the user to set a disparity level with which s/he is comfortable. This capability requires no physical adjustment of the actual camera or the camera mount, since it is accomplished simply by increasing or decreasing the number of frame between the L_image and R_image pointers. The interface also allows the user to save a frame to file if required.

---

4 NMEA messages are general ASCII based text data designed specifically for the use of navigation.
The software development process was one of the most time consuming portion of this thesis. It made use of a series of specialised video, graphics and desktop windowing libraries that was specific to the IRIX platform. A significant amount of time was invested in getting the different software functions to work in unison with one another, as well as with the required software. The code itself played an important part in the actual system. The single camera system was designed to minimise hardware requirement in implementation. It was also designed to reduce the number of expensive camera mounts needed, as well as to reduce alterations needed to the search aircraft (including minimisation of wiring and mounting locations). This afforded the system to be more adaptable to different type of search vehicles. However these simplifications to the hardware requirement meant that additional software development was necessary. Since the system hardware was not available at all times, it was necessary to make the software as robust as possible. For example, although the original intention was to use a NTSC video camera with 640x480 resolution, the final software and computer platform allows the use of a different camera, as well as various CCD resolutions. The software is also responsible for adjusting the time delay to generate the desired horizontal disparity.
3.4.4 Tradeoffs

Effect of Large Focal Length vs Effect of Vibration

By increasing its focal length, a video system can zoom to a small area of interest and magnify the image to enhance inspection. Since the image is magnified, any extraneous movement of the camera will also be increased. In most stationary systems, this is a minor problem. However, in an application in which the camera is mounted to an aircraft, small vibrations can cause the angular position of the camera to change by relatively large amounts. Furthermore, with longer focal lengths, small angular changes in the position of the camera will still result in relatively large movement of the video image on screen. Since the single camera system relies heavily on the stable camera mount, increases in focal length can present a serious problem.

Focal Length vs. Field of View (FOV)

One possible advantage of using any kind of a video system, be it stereoscopic or monoscopic, to supplement the current SAR search method is the potential flexibility of adjusting the zoom of the video image. By using lenses of different focal length, different magnifications of the video image can be achieved. The SAR technician would then not have to hold the binoculars in place for an extended period of time or to manage multiple pairs of binoculars at the same time. This capability can be beneficial in situations for which a low altitude pass is not possible. Furthermore, it can also be useful if the SAR technician decides to circle around for a second pass to examine a possible target. In such cases the magnification could be increased to allow better resolution to aid target confirmation. However, while the increase in focal length in this case would provide higher magnification, it would also do so at the expense of field of view (FOV). In other words, in order to see smaller targets in more detail, the total area captured on the screen will necessarily become smaller.

---

5 Using any kind of video system, it is also possible for the video to be recorded and replayed during flight, thus possibly obviating the necessity for some circling around manoeuvres. The video recording can also be processed digitally for zooming and/or with other enhancing techniques to improve detectability. It could also be possible for the inspection of potential targets to be carried out by another SAR technician without disturbing the operator who originally identified the potential target. In addition to the time savings, it is possible to get the opinion of a second, or even third, technician about a particular potential target.
Hyperstereo vs. Stereo FOV

By definition, with a hyperstereoscopic video system, the camera separation must be large in order to exaggerate the user's perceived depth. With the present single camera system, the set-up is equivalent to that of a dual camera stereo video system using cameras with parallel optical axes. As stated previously, this results in a reduced effective stereo FOV. Furthermore, as the effective camera separation increases, the effective stereo FOV of the system would continue to decrease in size.

This tradeoff can be closely related to the focal length as well. As the focal length of the camera lens increases, the magnification causes a reduction in the FOV. If the time delay between the left and right images remains constant, a reduced FOV obviously means a reduced stereo FOV as well. However, since the stereo image is software generated, it is possible to automatically compensate for changes in focal length as long as this information is available to the computer.

The ability of the software to compensate means that the reduction of FOV would have no effect on the overall hyperstereo configuration of the system. In other words, the effective camera configuration can simply be re-scaled to compensate for the change in focal length. (Similarly, the system can also easily re-scale for different search altitudes.) In a dual camera system, however, the camera separation would have to be changed physically to maintain usable hyperstereopsis in the event of a change in focal length.

3.5 Preliminary Field Evaluation: Kingston, Ontario

The original intention of the field trip to Kingston was to test the prototype system on a search vehicle. A parallel objective of the Department of National Defence was to evaluate an aerial reconnaissance camera system distributed by Wescam. A CH-146 Griffin helicopter was used as the aircraft for the evaluation. Unfortunately, the Griffin could not meet the AC power requirement needed to operate the Indigo 2 computer workstation on board. Furthermore, with the camera and recording equipment on board the helicopter, physical space was also a limitation. It was therefore decided that the video images would be recorded and the system tested by playing back the video tapes in the laboratory. For the purpose of initial evaluation, it was felt
that the effect of testing offline would be minimal. By using video tapes, the risk to damaging equipment was also reduced. With these constraints, the main objective of the trip to Kingston was thus to collect video data for the purpose of preliminary testing of the software.

The Kingston trip took place on Jan 19-20, 1998. As stated, the main purpose of the test flight was to evaluate a camera system offered by Wescam. Because the actual flight time available to gather data for this project was limited, the first problem created by this constraint was that the choice of terrain during the test flight was also very limited, to the vicinity of the airport. This made it difficult to collect video recordings that are representative of the type of terrain encountered during SAR operations. As a result, the video recordings obtained from the test are useful for demonstrating the concept of single camera stereo, but unfortunately not for demonstrating the feasibility of the system for SAR operations.

Several important observations were made, nevertheless, which were invaluable for the preliminary evaluation of the proposed system. Firstly, it was found that, with the Indigo 2 workstation used for this project, the video capturing speed was not sufficient for full speed video. This was because the amount of data to be manipulated was too large for the memory and bus structures of the workstation. It should be recalled that this was one of the anticipated problems with the single camera design.

Another observation made during the trip was that the Wescam camera system used during the test provided excellent vibration damping capability. It was thus concluded that the original requirement for vibration damping can in fact be met with existing camera mount technology. The Wescam system also provided the necessary degrees of freedom in camera orientation to overcome the possible problems associated with vertical disparity created by crosswind during flight. However, it should be noted that the camera correction was done manually by a camera operator. If such a single camera is to be adopted for actual SAR operations, this correction must be automated.

Finally, from the video data gathered during the trip, it was concluded that it is possible to generate stereo imagery using a single camera system. Although the testing was conducted in two parts (i.e., a recording was made and then fed into video input of the workstation at a later time), it was representative of actual operation of this type of system. The recording was made using a standard video recording protocol identical to the signal provided by the output of a video
camera. The fact that the video was not live thus made no difference with respect to the testing of the technical feasibility of the concept.
4 Experimental Evaluation & Performance Test

During the test in Kingston it was not possible to record useful video data to test the single camera system from the point of view of its potential efficacy for SAR operations. More specifically, the data obtained did not allow testing of the hypotheses that stereopsis can be used to de-camouflage hidden or hard-to-find targets, and that hyperstereopsis in particular, with its potential for enhanced depth resolution, can further improve performance. To test the system with respect to these hypotheses, different video data are needed. There are three major difficulties in locating the suitable type of video recording, however. Firstly, most of the existing video recordings from an aircraft do not fit the requirement of the proposed single camera stereo system, in terms of altitude, speed and challenging terrain. Secondly, it would not be economically feasible to record new data for testing purposes due to the expense involved in flight time and camera rental. Finally, even if the financial resources were available, there is virtually no way to control the environment, with respect to ensuring the presence of targets and controlling environmental factors such as weather, lighting and visibility conditions.

In order to solve these problems, it was decided that test stimuli would be created using computer simulation. Because digital simulation can generate video data which can be recorded onto a standard video tape, the hyperstereoscopic video system developed can be tested with standard NTSC video signals identical in form to the video recordings from the Kingston trip. Furthermore, because this allows complete control over all environmental factors, without the enormous expense associated with actual flight, the presence, location and features of targets can be guaranteed. It is also possible to artificially slow down airspeed to overcome the problem associated with the slow bus speed of the computer workstation.

Using computer simulation does, however, have some disadvantages, the most crucial of which is the lack of realism in the computer-generated image. Although it is possible to give some realism to the scene, by using different computer tools such as texture mapping, it should be noted that the resulting simulated stimuli are quite distinct from real video images taken with a video camera. Nevertheless, it should still be possible to use computer simulation as a means of validly testing the ability of the hyperstereoscopic camera system to improve basic detection performance in at least one kind of a simulated visual environment, even if that physical environment does differ significantly from an actual SAR scene. In other words, if the
experimental studies performed here do show advantages of the hyperstereoscopic video system, this does not necessarily mean that these advantages will transfer to real-world operations. That would still have to be confirmed empirically. However, were the converse to be true, that no advantages to the new system are to be found, this would have serious consequences as the desirability of transferring the technology beyond the laboratory.

Two different experiments were conducted to study the detection performance using the single camera hyperstereoscopic video system. A signal detection experiment was designed to study the potential benefit of hyperstereoscopic video in a simulated SAR-like detection task. A second experiment, based on the method of limits, was conducted to study the relationship between horizontal disparity and tolerance for vertical disparity.

### 4.1 Experiment: Visual Detection Using Hyperstereoscopic Video

The main motivation for this work was the possibility of using stereoscopic video technology as a tool to aid SAR technicians in detecting crash sites during aerial search. The experiment described here was designed accordingly, to test the theoretical advantages of hyperstereopsis with a prototype video system. Geometrically, a video system with a large camera separation should produce a larger horizontal disparity, i.e. "hyperstereo", in comparison with an "orthostereo" setting, when viewing the same scene. The resulting hyperstereo images should have a higher depth resolution, due to exaggerated relative depth. Perceptually, however, it is not possible to maintain fusion for binocular disparities of indefinite magnitude. That is, as disparity increases, maintaining fusion becomes increasingly difficult. The present experiment was thus designed to test not only the existence of the hypothetical advantages of using the hyperstereoscopic video system for detection, but also to investigate the limits of fusion with the system.

#### 4.1.1 Hypothesis

The main objective of the experiment was to evaluate the critical factors motivating the design of the system: that is, whether a (single camera) hyperstereoscopic video system could be of possible benefit for aerial SAR operations. Hypothetically, detection performance was expected to improve as horizontal disparity increases, in light of the fact that stereopsis has been
shown before to be of benefit for breaking camouflage and making masked targets more visible. Furthermore, strictly from geometrical differences, hyperstereopsis provides higher depth resolution and exaggerated depth differences. In detecting depth differences, hyperstereopsis may have some additional benefits as compared to regular stereopsis. However, it was also expected that, as the binocular (horizontal) disparity approaches the fusion limit of the observer, detection performance should degrade quickly, as the binocular fusion limit should interfere more and more with detection performance. This expected performance is illustrated in Figure 14. Since the fusion limit varies with individuals, the horizontal disparity for optimal performance may be different for each participant (Boff & Lincoln 1988 - Section 5.91).

![Figure 14 - Hypothetical detection performance as a function of horizontal disparity.](image)

4.1.2 Experimental Approach

During a SAR operation, the technicians must search over densely wooded terrain for cues to indicate the location of a crash site, the appearance of which can be very similar to its surrounding. Their task, in other words, is essentially one of detecting weak target signals in the
presence of noise. At the same time the SAR-techs must continually weigh the consequences of their decisions, such as whether to circle for another look or whether to request that a helicopter be dispatched to a candidate location for further, closer examination. The visual detection task of the SAR technicians thus has many similarities with traditional signal detection theory (SDT) tasks. For this reason, a laboratory experiment using signal detection theory was selected as the primary tool for measuring performance with the new camera system.

Although such experiments can yield informative results in laboratory settings, it is important to keep in mind that they are not generally representative of the actual tasks performed by the SAR technicians, who are typically required to monitor the surrounding terrain for targets continuously, for hours at a time. Nevertheless, although such differences between laboratory experiments and field studies are common, it was felt that the present experiment would still be valid from a basic perceptual point of view, rather than from the perspective of predicting accurate performance figures for actual SAR operations.

As in other SDT type experiments, the participants were asked to detect the presence of a designated stimulus during each trial. In traditional SDT experiments, however, the duration of each trial is predefined and is very short, usually along the order of hundreds of milliseconds to seconds. The nature of the system used for the present experiment dictates, however, that a moving scene must be used (due to the manner in which the stereo images are generated), in contrast to typically static images in conventional SDT experiments. At first glance, this difference may seem somewhat insignificant; however, it proved to be a major obstacle in the design of the experiment. On the one hand, continuous presentation did not present a problem regarding "trials" where a signal is present, since a hit (H) or a miss (M) could easily be registered based on the participant's response. It was also equally simple to identify a false alarm (FA), since the participant would have identified a target when none was present. It was the definition of a correct rejection (CR) which proved to be problematic, however, since the uncertainty about the duration of a single trial meant that a correct rejection was difficult to define. In such cases, essentially involving a null response, the participant would give no indication that an actual decision was being made about a particular region of the scene.

Green and Swets (1966) have described in their book many of the different possible applications of SDT to the estimation of various psychophysical parameters over a large range of
experiments. The material that is most pertinent to this study can be found in chapter 9 of that book, where Green and Swets describe a series of three experiments conducted by Egan, Greenberg & Schulman (1961). In that study Egan et al. were interested in validation of SDT in experiments where the duration of a trial and/or the duration of the stimulus can not be clearly defined, in other words cases involving the effects of temporal uncertainty about trial duration on experimental results. Their findings confirmed that temporal uncertainty does have an effect on detection performance. However, they also confirmed that their results in those experiments were compatible with those predicted in "conventional" signal detection theory. More specifically, they showed that, even with temporal uncertainty, the $d'$ parameter computed in SDT could still be used to rate sensitivity to the signal under different detection conditions. However, it was still necessary to carefully design a stimulus that would allow a reasonable definition of non-signal trials, so that CR's could be rated. The actual stimulus used for the experiment is discussed in the next section.

4.1.3 Design of Stimulus

In order to test detection performance with the new system, it was necessary to provide a video feed to the Indigo2 workstation. However, it was decided that, rather than using actual aerial video, the stimulus would be generated by computer simulation. This was accomplished by means of the Alias/Wavefront Power Animator™ software, a package which allows one to simulate visually very realistic physical worlds.

The stimulus was designed based on several different criteria. First of all, in order to maximise the likelihood of the results being accepted beyond the laboratory, it was decided to simulate as well as possible the essential elements of an aerial search and rescue operation over forested terrain. Secondly, on the other hand, it was necessary to keep the model used for the video scenes relatively simple, since the computational time required to generate higher fidelity video would have been too intensive to be practical, given the available resources. Thirdly, the detection difficulty level of the selected stimulus had to be such that floor and ceiling effects would not skew the results statistically.

After exploring several different models for the search terrain, a simple model was selected. As depicted in Figure 15, the terrain consisted of a flat surface with a multitude of
circular disks of two different heights scattered over it. These circular disks were intended to represent treetops. In order to simulate the foliage condition of the search environment, texture mapping was used to give the entire scene a green colour and a leaf-like appearance. Many different textures were tested before the selection was actually made. The final chosen texture represented the best compromise between realism and computational requirements. Even with the simple texture used, the time required to compute each 3 minute scene (at 800 frames/scene) was over 8 hours.

All target disks were placed at a height that was significantly above the surface, with the intention of representing an object protruding from the environment. All non-target disks had a height that was only slightly and imperceptibly above the actual flat surface. As discussed above, the definition of Correction Rejections (CR) was a problem that must be dealt with, and the purpose of the non-target disks was both to act as distractors and to assist in the definition of CR outcomes. By introducing the non-signal disks, it was possible to consider each distractor disk which was not (incorrectly) identified as a signal to be a correct rejection.

It is important to note that, unlike actual search and rescue tasks, where technicians would typically be searching for indentations in the forest cover, the participants in this experiment were asked to detect target that were protruding relative to the non-target. In spite of this departure, it was felt that the experimental results would still be useful for answering the fundamental research question concerning the possible benefits of hyperstereopsis in (aerial) detection.
What you will see on the experimental stimulus (in grey scale).

Figure 15 - Model used for experiment 1. (same figure was shown to participants to explain their task.)

Figure 15b - Actual appearance of the experimental stimulus (in grey scale).
Shown in Figure 15b is a portion of a single frame presented during one of the stimulus video scenes. Although the texture is presented here in grey, it should be noted that the actual texture used for the experiment was green in colour. In the image, both non-signal and signal disks are shown. It was designed to be difficult to see without motion.

Each run consisted of a three minute video clip shown to the participants. The number of targets within each run ranged from 2 to 6 in total. (This range was known to the participants, but not the actual number of targets for each trial.) As a result of the area of the simulated terrain and the simulated ground speed, each target was to be visible on the screen for a duration of 15 seconds. For the experiment, the ground speed was simply controlled by the animation and the effective camera separation was determined by precomputed time delay. During the 3 minute simulated fly pass, the participants were asked to detect and identify all possible targets.

4.1.4 Participants

Ten students were scheduled to participate in the experiment. After the initial two sessions, it was discovered that the highest disparity level had been set too low to cover the appropriate range of values, so it was increased and the experiment recommenced. Eight students (both undergraduate and graduate) from the University of Toronto participated in the modified experiment. Previous exposure to stereo display systems varied. A stereo acuity test was given to each participant to assess suitability for the experiment. Participants who were able to resolve a binocular disparity of 25 minutes of arc or less using the Randot Stereo Test kit (that is, difficulty level 8 out of 10 in the test kit) were accepted for the experiment. For participation in the 1.5 hour long experiment, each participant was paid fifteen dollars. To keep the experiment balanced, the data from the first two subjects were not used in the analysis presented here.

4.1.5 Experimental Design

The purpose of the experiment was to determine the benefits, if any, of hyperstereopsis in a simplified visual detection task. Furthermore, the experiment was designed to provide a quantitative measure of these benefits for different levels of horizontal disparity (HD). Five

\[ \text{Note that, with the actual operational system, which uses the GPS, the software adjusts separation and determines the required time delay.} \]
different levels of on-screen horizontal disparity were tested: 47.5, 95, 142.5, 190, and 285 minutes of arc. The stereoscopic video frame update rate was the same for all trials, approximately 1.5 Hz. The participants were seated at a constant distance of 50 cm from the screen. It should be noted that the monoscopic viewing condition was not evaluated in this experiment as a control condition. In fact, two possible approaches to including this condition were considered, but were rejected as impractical. First of all, the obvious choice of simply presenting the output of the stereoscopic video system with zero delay was considered. However, because of the nature of the stimulus in this case, it would have been effectively impossible for the participants to locate the target disks at the 1.5 Hz frame update rate. The second possibility was for the participants to view the video clips directly, that is, prior to the video software. However, viewing these clips directly would have provided a set of very different visual cues for target detection, since the texture of the target disks appeared to shift when played directly to the monitor at full frame rates. The target disks would therefore have been relatively easily identified, but due to a completely different set of non-stereoscopic characteristics. The comparison would thus not have been a fair one.

The experiment was a within-subjects design. Each participant was presented four different scenes at each of the five different disparity levels. In each video segment, there were between 2 and 6 targets. Along with these targets, there were approximately 80 non-target disks. The large number of non-target disks was intended also to introduce an element of time pressure, similar to that experienced by SAR technicians during actual operations. The participants were aware of the range of the number of targets. Also, they were informed that each trial would last approximately 3 minutes. The task was to scan the scene and identify all possible targets as quickly as possible. This was done verbally, by announcing the presence of the target while pointing at it.

4.1.6 Setup & Procedure

The session began with stereo acuity testing, as outlined above. Qualified participants were then given a short questionnaire regarding corrected vision and prior experience with stereo display systems. (See Appendix A for a sample questionnaire.) Following the questionnaire, the purpose of the experiment and the simulated task were explained to each participant with the
instruction handout found in Appendix A. At this point, participants were given the opportunity to ask questions about any aspects of the instructions which needed clarification.

At the beginning of the experiment, the participants were shown a sample signal that would appear on the screen, to ensure understanding of their task. During the experiment, a total of twenty trials, using 5 different HD levels, were given in random order. The order of presentation and the scenes used were randomised to minimise any possible learning effects. The participants were given a short rest period between each trial. In addition, the participants were told to notify the experimenter if more rest time was desired.

In each of the experimental trials, multiple targets (from 2 to 6) were presented. The participants were informed about the possible number of signals that could be presented during each video segment, but not the actual number. They were also told that it was possible to have more than one target on the screen at the same time. Their task was to identify targets by pointing directly at the screen. The participants were also asked to rate their confidence level on the certainty of their detection, using a scale from 1 to 5, where a rating of 5 indicated that he/she was sure that a target was correctly identified. A rating of 1, on the other hand, indicated that he/she was least sure, but suspected that a target was being shown on screen. The responses of each participant were recorded for analysis. Unlike a conventional SDT experiment, where an explicit "No" response to a non-signal target would constitute a Correct Rejection (CR), for this experiment a null response (i.e. participant says nothing) while a non-target disk was present was considered to be the equivalent of a CR.

4.1.7 Results & Discussions

In order to quantify detection performance at different horizontal separations, Hit (H) and False Alarm (FA) ratios were gathered during the experiment. This information is summarised here in Figures 16a and 16b, in the form of mean probabilities of hits and false alarms. Examining the basic detection performance data, it is clear that the participants were able to perform well in the detection task using the single camera system. The participants achieved reasonably good hit rates, ranging from a low of 54% to a high of 75% (Figure 16a), with only a small proportion of false alarms, ranging from a low of 5% to a high of only 8% (Figure 16b).
Using the H and FA data, the d' values were calculated for each participant at each of the different HD levels. In simple terms, the d' values were taken as a measure of the sensitivity of the participant to the experimental stimulus. The following set of figures summarise the results of the experiment, in terms of mean d' (Figure 17 and Figure 19), as well as individual performances (Figure 18 and Figure 20) and receiver-operating-characteristic (ROC) curves (Figure 21).

The results obtained during the experiment for the relationship of d' as a function of HD were not as unambiguous as was originally hoped. The predicted performance characteristics were not found. Figure 17 shows the average d' values, with standard deviations, at each of the different horizontal disparity (HD) levels tested. The results indicate that performance for the first 5 HD levels appears not to change, with perhaps only a slight dip in performance at the highest HD level. A one-way analysis of variance showed that there was no significant difference among the d' levels due to the different levels of HD ($F(4, 35) = 1.61, p > 0.1$).  

---

Once again, it should be recalled that the first two participants' data were removed for the reason explained earlier.
Figure 17 - Average Detection Performance (d') for all 10 participants.
Figure 18 - Detection performance (d') for individual participants.
It is obvious that the shape of the performance curve did not coincide with the hypothetical U-shaped performance curve illustrated in Figure 14. To gain a better understanding of the experimental results, it was deemed necessary to analyse the data for each individual participant more closely. The d' data from each of the participants were thus plotted and presented in Figure 18. Examining these plots, it seemed possible that the absence of the hypothesised behaviour in the mean d' plot could be explained by individual differences, since peak detection performance was spread across the four lower HD levels for different participants, rather than occurring at a single identifiable level of HD. In other words, since the best performance appeared to have occurred at a different HD level for each participant, this caused the mean d' over all subjects to be similar at each level of HD, even though an optimal HD level might in fact exist for each individual. Furthermore, although each of the participants was tested for stereo acuity against criterion, their individual sensitivities to the signal could have been quite
different. Finally, most of the participants found it difficult to perform the detection task for the largest level of HD. Many of them commented on seeing "double", indicating the loss of fusion that was expected when HD became too large. This should have caused a degradation of performance for large HD, as expected, at least for a majority of the participants. Although this decline failed to achieve significance, Figure 17 does show a slight indication of this tendency.
Figure 20 - Individual subjects' detection performance ($d'$) with confidence rating of 2 and above being considered as a "yes" response.
Figure 21 - ROC curves for individual subjects.
Since confidence ratings were used during the experiment, another set of d' plots was generated for a different decision criterion, by considering all ratings of 2 or above to be a "yes" response. In other words, all "1" responses, which previously would have been tallied together with the "yes" responses, were now considered as "no" responses. This adjustment of the cutoff level reduced not only the number of Hits, but also the number of False Alarms, for all "1" responses. (The data were such that it was not possible to analyse them further at a higher cutoff levels, since this would have resulted in FA rates which approached zero.)

The mean d' value with confidence cutoff of 2 was plotted, in Figure 19, which shows the general shape of the original graph in Figure 17. This was not entirely unexpected because, if individual differences (i.e. different peak values for each participant) were exerting their influence, the overall effect would likely remain, regardless of the confidence rating cut off.

Comparing the individual graphs in Figure 18 and 20, however, a peak became somewhat more apparent in many of the participants' d' performances. To test the possibility that individual differences were masking a weaker effect of the varying HD level, a second anova was performed. Instead of treating all data as equal, however, the data were analysed again using a randomised blocks design. Randomised blocks designs are often used in experiments where two different kind of effects, those of the treatments (the HD level) and those of the blocks (participants), are at work. The interested reader should consult other statistic references for details (e.g. Box, Hunter & Hunter 1978). The result of the second analysis showed that there was in fact a significant effect due to individual differences ($F(7, 28) = 148$, $p < 0.01$).
Furthermore, HD levels now showed a significant effect on the participants’ detection performance \( (F(4, 28) = 4.45, p < 0.01) \). The actual Anova tables can be found in Appendix B.

In an attempt to better understand the data, *receiver-operating-characteristic* (ROC) curves were plotted for each of the participants. This was done by varying the cut off levels of the confidence rating and computing the resulting probability of hits and probability of false alarms for each level [Anderson, D. C. & Borkowski, J. G. 1978]. The resulting plots, shown in Figure 21, summarise the performance of each participant for each level of HD. (A larger version of these plots can be found in Appendix C). In theory, a ROC curve which runs along the top-right to bottom-left diagonal of the plot would indicate chance performance, whereas perfect performance would be indicated by a ROC curve which was closest to the top-left corner of the plot. The main advantage of the ROC curves is that they eliminate the need to evaluate d’ values at each of the confidence rating levels. Furthermore, because the area under the ROC curve for each of the viewing conditions can be used as an index of detection performance for that individual for that particular HD level, this measure can be used as a means of gauging performance in addition to d’.

The area under the curves gave the expected probability of correct detection at each HD level. The mean areas under the curves were calculated across the different participants (Figure 22), and a randomised blocks design anova showed similar results to those observed on the d’ measure \( (F(4, 28) = 5.484, p < 0.01) \). The results found in this analysis thus corroborated those found during the d’ analysis. The slight decrease in performance at the highest horizontal disparity is still present, and the variability between individual participants is equally high. It should be noted that the expected performance (represented by the area under the ROC curve) is quite high at approximately 0.8. Unfortunately, it was not possible to compare this performance datum to an equivalent monoscopic task because of the design of the experimental stimulus (see Section 4.1.5).

---

8 The interested reader should refer to Green & Swets (p.42) for an explanation of the actual procedure with which the ROC curves were generated and for the relationship between the area under the curve and detection performance.
The original objective of the experiment was to identify whether hyperstereopsis can maximise detection performance. The experimental results showed that the hypothesised performance was found only in some of the participants. For others, hyperstereopsis may have been disruptive. Furthermore, detection performance varied greatly from individual to individual, which may have been the result of individual differences in adapting to different horizontal disparities. For this experiment, finer resolution in the adjustment of HD level was unfortunately not possible due to the hardware constraints. That is, using the test platform, a higher video capture rate was not possible, so, in order to obtain finer resolution in effective camera separation, the time between consecutive frame capture moments must be reduced. Consequently, for those participants whose best performance occurred at a HD level of 49.5 min of arc, it was possible that the left hand side of the expected inverted U-shape performance (Fig. 13) was simply not captured due to this technological constraint on the lowest setting.
4.2 Performance Test: Vertical Disparity Tolerance

One of the key concerns identified early in the design of the single camera hyperstereoscopic video system was the effect of crosswinds on the quality and usefulness of the final stereo video image. Just as the forward motion of the aircraft is used to create horizontal disparity, any lateral motions, such as those due to crosswinds for example, if left uncorrected via counter rotations of the camera, will result in vertical disparity. Vertical disparity refers to the situation in which the right and left eye images of a stereo pair are vertically out of alignment. Although vertical disparity has been found to be a generally disruptive effect in binocular viewing, observers can nevertheless tolerate small amounts of vertical misalignment while maintaining binocular fusion. As the amount of vertical disparity increases, however, it becomes increasing difficult to maintain fusion. It has been reported that tolerance for vertical disparity is approximately 25 minutes of arc under controlled viewing conditions (Engineering Data Compendium, 1988 Section 5.906). An understanding of tolerance for vertical disparity can thus provide important insights for actual implementation of the present system in a SAR environment.

The performance test reported in this section was conducted in collaboration with Rishi Bageshwar, a fourth year student in the Industrial Engineering programme at the University of Toronto, within the context of his fourth year thesis project [Bageshwar 1999]. Mr. Bageshwar participated in the design of the test and was responsible for the gathering of the data.

4.2.0 Objectives

The main objective of the experiment was not to replicate the literature on vertical disparity tolerances but to explore the particular effect of horizontal disparity on the vertical disparity tolerance of observers within the context of the single-camera stereo system. This has important implications for the hyperstereoscopic video system since it relies on exaggeration of relative depth to increase the likelihood of target detection. Although increased disparity can magnify depth differences, it may also decrease the stability of binocular fusion. It is therefore reasonable to expect that increases in horizontal disparity will result in some decrease in tolerance for vertical disparity (in addition to eventual diplopia due to exaggerated horizontal
disparity). The expected results for the present performance test was that, as horizontal disparity increases, users' tolerance of vertical disparity for maintaining fusion will be reduced.

4.2.1 Performance Test Method

In order to study tolerance to vertical disparity, the method of limits was selected for the test. For additional details on the method of limits, interested readers can consult Anderson & Borkowski (1978) for more information on Fechner's original works "Elements of Psychophysics". In a typical method of limits experiment, participants are required to make judgements regarding the state of a stimulus which is continuously changing in one direction or another. For example, in an experiment designed to explore visual acuity, a participant may be asked to judge the relative length of two lines. Two lines are presented, one with a fixed length while the length of the second line is varied during the experiment. At the onset of the trial, the line with variable length (stimulus) will be either longer or shorter than the comparison line with fixed length (standard). The participant will be asked to compare the length of the stimulus to the standard, and judge their relative lengths. If the length of the stimulus at the beginning of the trial is shorter than that of the standard and the participant responds correctly, the length of the stimulus will then increase and the participant will make another judgement. This continues until the participant identifies the lines as being equal in length. The trial will end when he/she perceives the length of the stimulus to surpass that of the standard. Conversely, if the stimulus begins with a length longer than that of the standard, the adjustments will be in the other direction and the response sequence will be reversed. In this case, the trial will end when the participant indicates that the stimulus is shorter than the standard.

In this performance test, the tolerance for vertical disparity (VD) was investigated using a similar method of limits experiment. Like the example experiment, the vertical disparity test consisted of trials of both increasing and decreasing values of VD. This consisted of a simulated flight over a flat surface. For each overflight, a different horizontal disparity was introduced by the single camera stereo software. During the performance test, the vertical disparity either gradually increased or decreased. The participant was asked to concentrate on the stimulus and signal the point in time at which diplopia (or loss of fusion) occurs. In trials with increasing VD, the vertical disparity between the left and right image was small at the onset and increased progressively. The participant typically would perceive a single fused image at the beginning of
the trial, with fusion usually (but not always) being lost as the trial proceeded and VD increased. During a decreasing VD trial, the initial disparity was such that the participant usually perceived a double (unfused) image of the target at the beginning of the trial and then reported when fusion was achieved as the trial progressed.

**Hypothetical Results:**

- ◆ Point of loss of fusion in increasing VD trials.
- ● Point of fusion in decreasing VD trials.

![Graph](image)

Figure 23 - Expected (hypothetical) results for experiment 2. (This type of hysteresis characteristic is typical of method of limits experiment.)

Similar to other method of limits experiments, it was expected that thresholds would be direction dependent, that is, that the value of VD at which loss of fusion occurred during increasing trials would be different than that at which fusion is achieved during decreasing trials. More specifically, it was expected that fusion would be maintained for a larger VD range in trials where the vertical disparity was small at the onset of the trial and increased with time. Conversely, it was expected that the participant will achieve fusion at a smaller VD than the point where fusion was lost in the increasing VD trials. The hypothetical results for an experiment such as that described above are illustrated in Figure 23.

### 4.2.2 Stimulus

To study tolerance for vertical disparity (VD) in the single camera hyperstereoscopic video system, a special stimulus was created using Alias Power Animator™ software. Since detection was not the focus of this particular study, the stimulus was designed to be easily discernible from the background, under quasi-optimal viewing conditions. The target was a cross shaped column, illustrated in Figure 24, rendered in blue and raised from a flat green background.
surface. Because, as described above, cross wind is an anticipated disturbance, which, if left uncorrected, can cause vertical disparity in stereo pairs during actual SAR flights, it was originally decided to simulate a search vehicle travelling in a straight line path, with a constant crosswind. Vertical disparity arises in this case because the path is not perpendicular to the optical axis of the CCD, as illustrated in Figure 10. For such a case, however, this would result in a constant VD, which would thus not be suitable for the planned method of limits experiment. Instead, rather than simulating a constant crosswind, animation was used to generate a curved fly pass trajectory, as shown in Figure 24. This path thus created a situation where the VD of the stereo pair was either increasing or decreasing, depending on whether the direction in which the path was traversed was respectively from right to left or left to right, as shown in the figure.

![Figure 24 - Experimental stimulus for Experiment 2. By moving the cross-shaped target along the curved path, vertical disparity will gradually increase (right to left) or decrease (left to right).](image)

4.2.3 Participants

The nine participants in the test were undergraduate engineering students at the University of Toronto. Before the actual performance test, each participant was tested for stereo acuity as described in Experiment 1. Participants who failed to meet the stereo acuity criterion were not permitted to participate in the actual experiment. All individuals who participated in the experiment had little or no prior experience with stereo display systems.

4.2.4 Performance Test Design

The test was conducted using the same animated fly pass sequence as stimulus in all cases. (It should be noted that the signal simulated by computer was identical to that of a real
video camera, i.e. NTSC). The second independent variable of the test was the time delay used to generate stereo video for the single "camera". For this test, four different time delays, corresponding to four different horizontal disparities, were used. These were obtained by controlling the number of frames separating the two stored images used to display left and right eye views to the participants. By setting the frame separations to 0, 1, 2 and 3, four different time delays were tested. To clarify, in the 0 frame separation condition, two consecutive frames were presented as a stereopair.

The test followed a within subjects design with two trials for each time delay. It should be noted that during the performance test, the horizontal disparity was not precisely controlled (see figure 24). This was the result of two different contributing factors that were technically difficult to resolve. Firstly, as discussed earlier, due to slow throughput of the computer system bus, the resolution for adjusting HD was very low. The update rate of the stereoscopic video system consequently was approximately 1.5 Hz for all trials. During the test sessions, it was necessary to use four settings incremented by the smallest resolution available with this system (a single video frame). The second factor which caused the "wandering" HD levels was due to an error made during the animation of the stimulus video. In the animation process, the speed at which the camera traversed the "flight path" was constant. However, the speed was held constant along the path instead of in the horizontal direction. Therefore, as the vertical component of the speed increased, the horizontal component would be reduced. This resulted in the inability to control HD at a constant level for each flypass, since the system's resolution was too coarse to allow further corrections. Fortunately, even without the precise control of HD, the time delays were sufficiently different to allow the resulting data to be clustered in four different areas along the horizontal axis. These time delay settings also covered a wide enough range to provide information about a useful range of operating condition.

4.2.5 Setup & Procedure

The study was carried out on the same SGI Indigo 2 workstation used in the software development. Participants' viewing distance was fixed at 50 cm using a chin rest. At the beginning of the test, participants were given written instructions about the performance test. The instructions included an illustration of the target stimulus, showing both the fused and unfused state of the target to ensure that the participants understood the differences. The
participants were told that there were two different types of trials, starting with either a fused or an unfused target stimulus. The participants were instructed to follow the target as it traversed the screen and indicate verbally when a change of state in the target occurred (i.e. fused to unfused, or unfused to fused). The video scene was played back on a separate computer workstation. In essence, the second computer acted as a camera. The main advantage here is the improved control over the video playback. Whenever the participant indicated a change of target state, the time index of the animation playback was recorded. This was then subsequently converted into actual horizontal and vertical disparities for the data analysis.

At the beginning of an increasing VD trial, the stereo pair was presented with zero vertical disparity (VD). As the target travelled across the screen the vertical disparity began to increase, due to the nature of the curved path being flown (Figure 23). The participants were asked to keep their head fixed using the chin rest, to ensure that they would not compensate for the increasing VD by tilting their heads in the same direction as the curved path. The participants were also asked to verbally indicate loss of fusion to the experimenter by simply saying "now". As the target continued its pass, the VD continued to increase. At the end of the pass, the direction of travel of the stimulus would reverse, and the VD of the stimulus would decrease as it traversed the screen (from left to right). The participants were again asked to verbally indicate when they succeeded to fuse the image. Each time the target traversed its loop (travelling from right to left on the screen then back to the right, following the exact path in the reverse direction) was considered a single trial of the test. Each participant performed a total of eight such trials (four different time delays presented twice each). Between each trial of two passes, the participants were given rest time until they were comfortable to continue.

4.2.6 Results & Discussions

The results collected from all nine participants are summarised in Figure 25 in the form of a scatter plot. It should be noted that during eight of the trials with increasing VD, i.e. right to left traversal where the image was fused at the onset, fusion was maintained (that is, participants failed to report loss of fusion) throughout the entire trial. Conversely, during thirteen trials with decreasing VD, i.e. from left to right, fusion was never achieved. These data are represented with up and down arrows respectively in the figure, showing the number of times that this occurred, for each value of HD.
In order to translate the time index of the video scene at the time the fusion state changed to the appropriate VD and HD values, a manually intensive process was carried out. Firstly, by knowing the time delay that is represented by a single frame in the memory buffer, it was possible to obtain these values through interpolation of two custom graphs (see Appendix D). These graphs were created by measuring the actual disparities (VD & HD) between consecutive frames at different points in the video clip. These data were then plotted, making interpolation of the VD & HD values possible.

Although horizontal disparity values are distributed widely across the horizontal axis, the grouping of points according to the four values of time delay is fairly obvious. This was the result of the two factors described in section 4.2.5. These factors also presented a problem during the analysis of the data. Because the HD levels could not be precisely controlled, an ANOVA type analysis was not possible. However, the scatter plot can still provide some very useful guideline information. For example, if HD were set to be above 100 minutes of arc, fusion would have been questionable for ten of the 45 observed trials (i.e. data to the right of 100 min of arc on the HD axis) at as little as 10 minutes of arc of VD (22%). Under the same HD condition, 24 of the 45 observed trial trials were below a VD value of 20 minutes of arc (53%).

In conclusion, the data gathered from this performance test provided an indication of the order of magnitude of the region of tolerated vertical disparities for different HD levels for the single camera stereoscopic system. Consequently, the results provide empirical support for the hypothesised prediction that tolerance of vertical disparities during hyperstereoscopic binocular fusion becomes less stable with concurrent increases in horizontal disparity.
Figure 25 - Tolerance of Vertical Disparity
5 Conclusions

The main focus of this thesis was a study of the potential usefulness of hyperstereopsis for aerial search and rescue (SAR) and reconnaissance. To accomplish this, a short study of SAR operations was conducted and a hyperstereoscopic video system was designed and prototyped. Then, using the prototype software, two different psychophysical studies were conducted to evaluate some of the human factors issues associated with hyperstereopsis and its possible application to SAR.

In aerial SAR, one of the key means used to identify possible crash sites is visually detecting broken treetops within the foliage. When searching strictly by eye, the key visual cues available are shadows and occlusion. It was proposed that the introduction of a stereoscopic video system could help the SAR technicians in their detection task by providing binocular disparity as an additional (optional) depth cue. It was also hypothesised that hyperstereopsis, that is, exaggerated binocular disparity, would permit better than "normal" stereoscopic viewing performance. Geometrically, it was reasoned that a larger separation (or effective separation) of the camera pair would provide improved depth resolution since optically increased binocular disparity exaggerates the relative depth differences between objects.

A single camera system was designed and prototyped. Two laboratory studies were performed, one to test the theoretical benefits and the other to examine some of the associated shortcomings. The conclusions from these studies and their implications are summarised and discussed in the present section.

5.1 Experiment 1: Hyperstereopsis

The following points summarise the findings of Experiment 1:

1. The hypothesised primary advantage of hyperstereopsis for enhanced target detection was not clearly established.

2. Optimal horizontal disparity is apparently highly dependent on individual users, and can vary across a wide range of values.
During the design of the hyperstereopsis experiment, the key concern was the actual selection of the stimulus. Since this was designed to be a signal detection type of experiment, it was important to ensure that the signal and noise states were sufficiently similar to eliminate possible ceiling effects, i.e. perfect detection performance. However, due to the computational requirements involved, the actual choice of stimuli was limited in terms of realism and complexity. At the same time, it is equally important to avoid floor effects, i.e. where the task is so difficult that performance is equally poor for all conditions. If a floor effect had been at work during the experiment, the advantage for hyperstereopsis might have been better illustrated with a more difficult detection target. As a consequence, even with the extensive effort spent on designing the stimulus, it was not practically possible to guarantee that the result produced would be able to capture the effect of hyperstereosis.

There was evidence which suggested that different operators are likely to have different "optimal" horizontal separations for this kind of detection task. Geometrically, a larger horizontal separation between the two viewpoints should yield exaggerated depth differences, and thus improved performance for (detection) tasks requiring high depth resolution, but only within the fusion limits of individual observers. Since this fusion limit appeared to vary with individuals, it is possible that each participant might have exhibited peak performance at a different level of HD. If this were the case, this would imply that we should not necessarily reject the hypothesis, but that it could be better tested if the system hardware were to allow for finer control of HD level. One of the main practical implications of the experiment, therefore, is that, due to individual differences, it may be important to provide flexible control of horizontal disparity to the SAR technicians.

5.2 Performance Study: Vertical Disparity Tolerance

The following points summarise the findings of Vertical Disparity Tolerance performance test:

1. Tolerance for vertical disparity was reduced in hyperstereoscopic viewing conditions as the magnitude of binocular disparity increases. In other words, at low horizontal disparity, fusion is relatively stable and a larger vertical disparity can be tolerated. However, at large horizontal disparities, along the order of 100 minutes of arc or higher, the rapid reduction of tolerance indicates a large reduction of binocular fusion stability.
2. The hysteresis effect, which can be found in many method of limits experiments, was observed from all test participants.

In general, the performance test corroborated the hypothesis regarding the expected reduction of stability of binocular fusion with increased horizontal disparity. Vertical disparity was identified early on as one of the potentially most significant factors affecting perceptual performance with the system due to the fact that, in aerial SAR operations, crosswinds can cause the airplane to fly in a direction relative to the ground that is not aligned with the fuselage. Although there are in principle camera mounts available which can correct this kind of problem, it is still very useful to know the precision required for realizing such corrections. From the results, it was found, for example, that when operating with horizontal disparities of more than 100 minutes of arc, tolerance of vertical disparities falls off fairly rapidly and VD should thus reasonably be kept to below 10 minutes of arc. Such a conservative estimate would allow for a wide range of both users and operating conditions.
6 Recommendations

The project reported here comprised essentially two parts: the development of a prototype of a single camera stereoscopic video system and a series of experimental studies based on this system. According to this division, as discussed below, two different types of recommendations for future research have been made from this thesis: research to improve the performance of the system and research to enrich our understanding of hyperstereopsis as a practical tool.

6.1 Recommended Technical Advances

Incorporation of camera control into video system.
Currently, the stereoscopic video system has no control over the camera mount. By allowing such control functions, it is possible for the system to correct for crosswind using GPS data.

Modification of software to execute on more advanced workstations.
This will reduce some of the limitations in resolution of horizontal disparity. Furthermore, this might also increase the display frame rate of the system.

Modification of software to communicate with aircraft navigational GPS.
Using a military GPS system, the precision in location and/or ground speed measurement needed for maintaining a constant HD can be greatly improved. This should reduce the chances of losing binocular fusion during an operation.

Incorporation of image processing algorithms to cope with changes in flight altitude during search operations.
Changes in flight altitude will cause changes in object size over time. If such changes are small, this will not be a problem perceptually. However, for larger changes, it is possible to provide appropriate corrections computationally.
6.2 Recommended Experimental Work

Reduce the strength of the stimulus used in Experiment 1.

Reducing the signal strength could better test the sensitivity of the system and thus provide a better test of the original hypothesis regarding the effects of horizontal disparity. Increasing the number of data (via additional participants and/or trials) may also be a key factor for eliminating the effects of individual differences.

Investigate the effects of camera angle (with respect to the vertical) on subjective comfort levels and on detection performance.

Another parameter that can be altered in the system is the direction in which the camera is pointed. For Experiment 1, it was pointed vertically down from the simulated flight path. It is also possible, however, to set the camera at a different angle on the plane that is perpendicular to the flight direction as described in figure 12, looking outwards rather than downwards.
7 References


Appendix A - Experiment 1 - Supplementary Material

Experiment: Hyperstereopsis in Simulated SAR visual search

Objective: The main objective of the experiment is to study the effect of a new method of visual presentation on a simulated Search and Rescue (SAR) task.

Task: During the experiment, you will be asked to wear a pair of stereoscopic glasses and view a series of video clips in 3D. Each video clip is designed to simulate a number of camouflaged targets which you are to detect while flying over a forested terrain and looking down at it. Your task within this period is to identify targets by pointing to them on the screen. Upon identifying a target, you will be asked to rate how certain you are that it is present by giving a confidence level, using a scale from 1 to 5 (a rating of "1" being least certain, and a rating of "5" being very sure). The complete experiment will consist of 20 individual trials of 2 minutes each. In between each trial, a rest period will be given. If additional rest time is desired, simply indicate this to the experimenter between trials.

Terrain Model: A simple terrain model was used for the video clip used in the experiment. The ground is represented by a simple flat surface with a green texture. Two different types of round disks (with the same green texture as the background surface) are scattered around this surface. The first type was placed at the same height as the flat surface. These are non-target disks and there are many of them in the scene. The second type was placed significantly above the background. These are the target disks that you need to identify (see figure below). There are between 2 to 6 of these target disks in each 2-minute video clip. It is possible to have more than one target on screen at the same time.

Figure:
Informed Consent by Subjects to Participate in Research Project

"Target Detection using Stereoscopic Video"

Investigator: Kit Man Cheung

I hereby consent to participate in a research project experiment "Target Detection using Stereoscopic Video". I understand that this approximately 90 minutes experiment that will include the Randot Stereotest for testing my stereo-acuity, a series of perceptual tasks which has been explained to me, and a questionnaire. In addition, with the understanding that my identity will in no way be divulged to anyone other than the investigators, I consent to have the results of my participation in the experiment published as part of this research. I understand that I will be reimbursed $15 for my participation at the conclusion of the experiment. I will have the right to withdraw from this experiment at any point of this experiment without any penalty, and to request that my data be destroyed.

I understand the procedures to be used on this experiment and the personal risks to me in taking part.
I understand that I may withdraw my participation in this experiment as any time.

Participant’s Name:

Participant’s Signature:

Date:
Demographic Questionnaire

For participation in the experiment “Orientation and Navigation in Minimal Access Environments”

Subject No.

Date:

Time:

Randot Stereotest Result:

1. Age: <21 21-30 31-40 41-50 51-60 >60 (Please circle one)

2. Gender: M F (Please circle one)

3. Do you wear corrective lenses of any kind? (Please circle one)
   Yes (please also specify): ____________________________
   No

   If yes, are you wearing prescribed lenses or contact lenses right now?
   Yes
   No

4. Prior experience viewing with a stereoscopic system? (Please circle one)
   Yes
   No

   If yes, how many hours?
   _____ < 5 hours
   _____ 5- 50 hours
   _____ > 50 hours
### Appendix B - Experiment 1: Anova

#### ANOVA: Simple One-way design

<table>
<thead>
<tr>
<th>Subject</th>
<th>HD Level:</th>
<th>level 1</th>
<th>level 2</th>
<th>level 3</th>
<th>level 4</th>
<th>level 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND</td>
<td>2.778</td>
<td>3.388</td>
<td>3.399</td>
<td>3.385</td>
<td>3.25</td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>2.134</td>
<td>1.781</td>
<td>1.315</td>
<td>1.737</td>
<td>1.209</td>
<td></td>
</tr>
<tr>
<td>SH</td>
<td>2.231</td>
<td>2.298</td>
<td>2.266</td>
<td>2.205</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>JH</td>
<td>3.031</td>
<td>2.264</td>
<td>2.111</td>
<td>2.283</td>
<td>2.009</td>
<td></td>
</tr>
<tr>
<td>DM</td>
<td>1.901</td>
<td>1.344</td>
<td>1.145</td>
<td>1.565</td>
<td>1.47</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>1.89</td>
<td>2.669</td>
<td>2.138</td>
<td>2.681</td>
<td>1.75</td>
<td></td>
</tr>
<tr>
<td>CCh</td>
<td>3.135</td>
<td>1.538</td>
<td>1.761</td>
<td>2.203</td>
<td>1.489</td>
<td></td>
</tr>
<tr>
<td>GJ</td>
<td>3.062</td>
<td>2.113</td>
<td>1.688</td>
<td>1.939</td>
<td>1.895</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>dof</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD Level</td>
<td>2.506601</td>
<td>4</td>
<td>1.609409</td>
<td>&gt;0.25</td>
</tr>
<tr>
<td>Residual</td>
<td>13.62783</td>
<td>35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### ANOVA: Randomised Blocks Design

<table>
<thead>
<tr>
<th>Subject</th>
<th>HD Level:</th>
<th>level 1</th>
<th>level 2</th>
<th>level 3</th>
<th>level 4</th>
<th>level 5</th>
<th>Subject Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND</td>
<td>2.778</td>
<td>3.388</td>
<td>3.399</td>
<td>3.385</td>
<td>3.25</td>
<td></td>
<td>3.228</td>
</tr>
<tr>
<td>MD</td>
<td>2.134</td>
<td>1.781</td>
<td>1.315</td>
<td>1.737</td>
<td>1.209</td>
<td></td>
<td>1.6352</td>
</tr>
<tr>
<td>SH</td>
<td>2.231</td>
<td>2.298</td>
<td>2.266</td>
<td>2.205</td>
<td>1.2</td>
<td></td>
<td>2.04</td>
</tr>
<tr>
<td>JH</td>
<td>3.031</td>
<td>2.264</td>
<td>2.111</td>
<td>2.283</td>
<td>2.009</td>
<td></td>
<td>2.3396</td>
</tr>
<tr>
<td>DM</td>
<td>1.901</td>
<td>1.344</td>
<td>1.145</td>
<td>1.565</td>
<td>1.47</td>
<td></td>
<td>1.485</td>
</tr>
<tr>
<td>CC</td>
<td>1.89</td>
<td>2.669</td>
<td>2.138</td>
<td>2.681</td>
<td>1.75</td>
<td></td>
<td>2.2256</td>
</tr>
<tr>
<td>CCh</td>
<td>3.135</td>
<td>1.538</td>
<td>1.761</td>
<td>2.203</td>
<td>1.489</td>
<td></td>
<td>2.0252</td>
</tr>
<tr>
<td>GJ</td>
<td>3.062</td>
<td>2.113</td>
<td>1.688</td>
<td>1.939</td>
<td>1.895</td>
<td></td>
<td>2.1394</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>dof</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>146.6524</td>
<td>7</td>
<td>148.956</td>
<td>0.000**</td>
</tr>
<tr>
<td>HD Level</td>
<td>2.506601</td>
<td>4</td>
<td>4.455454</td>
<td>0.007</td>
</tr>
<tr>
<td>Residual</td>
<td>3.93814</td>
<td>28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix C - Experiment 1: ROC curves
Area under the ROC curves for each participant
(Best performance for each individual is highlighted)

<table>
<thead>
<tr>
<th>HD Level</th>
<th>CC</th>
<th>CCh</th>
<th>DM</th>
<th>GJ</th>
<th>JH</th>
</tr>
</thead>
<tbody>
<tr>
<td>47.5</td>
<td>0.796407</td>
<td>0.878925</td>
<td>0.833579</td>
<td>0.879975</td>
<td>0.931306</td>
</tr>
<tr>
<td>95</td>
<td>0.900323</td>
<td>0.697245</td>
<td>0.764663</td>
<td>0.768566</td>
<td>0.853875</td>
</tr>
<tr>
<td>142</td>
<td>0.853884</td>
<td>0.719822</td>
<td>0.700132</td>
<td>0.694524</td>
<td>0.83046</td>
</tr>
<tr>
<td>189.5</td>
<td>0.877632</td>
<td>0.821817</td>
<td>0.788754</td>
<td>0.769014</td>
<td>0.865082</td>
</tr>
<tr>
<td>237</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>284</td>
<td>0.774538</td>
<td>0.672243</td>
<td>0.768832</td>
<td>0.776991</td>
<td>0.815763</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HD Level</th>
<th>MD</th>
<th>ND</th>
<th>SH</th>
<th>WW</th>
<th>MH</th>
</tr>
</thead>
<tbody>
<tr>
<td>47.5</td>
<td>0.829044</td>
<td>0.798454</td>
<td>0.902651</td>
<td>0.793245</td>
<td>0.886863</td>
</tr>
<tr>
<td>95</td>
<td>0.781672</td>
<td>0.931434</td>
<td>0.912059</td>
<td>0.816345</td>
<td>0.992837</td>
</tr>
<tr>
<td>142</td>
<td>0.567745</td>
<td>0.907514</td>
<td>0.884085</td>
<td>0.853639</td>
<td>0.915919</td>
</tr>
<tr>
<td>189.5</td>
<td>0.731318</td>
<td>0.937677</td>
<td>0.917111</td>
<td>0.865829</td>
<td>0.72341</td>
</tr>
<tr>
<td>237</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>284</td>
<td>0.597606</td>
<td>0.774425</td>
<td>0.741143</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Subject JH: ROC

Subject MD: ROC
Subject MH: ROC

Subject ND: ROC
Appendix D - Experiment 2: Interpolation Graphs

Experiment 2: finding VD & HD values through interpolation
Horizontal Disparity vs Time

- Test 1
- Test 2
- Test 3
- Test 4