The Design, Development And Evaluation Of An Active Stereoscopic Telepresence System

by

Richard Asbery

Thesis submitted for the degree of Doctor of Philosophy

March 1997
ABSTRACT
The work presented in this thesis documents the design, development and evaluation of a high performance stereoscopic telepresence system. Such a system offers the ability to enhance the operator perception of a remote and potentially hazardous environment as an aid to performing a remote task. To achieve this sensation of presence demands the design of a highly responsive remote camera system.

A high performance stereo platform has been designed which utilises state-of-the-art cameras, servo drives and gearboxes. It possesses four degrees of freedom; pan, elevation and two camera vergence motions, all of which are controlled simultaneously in real-time by an open architecture controller. This has been developed on a PC/AT bus architecture and utilises a PID control regime. The controller can be easily interfaced to a range of input devices such as electromagnetic head tracking systems which provide the trajectory data for controlling the remote mechatronic platform.

Experiments have been performed to evaluate both the mechatronic system and operator oriented performance aspects of the telepresence system. The mechatronic system investigations identify the overall system latency to be 80ms, which is considerably less than other current systems. The operator oriented evaluation demonstrates the necessity for a head tracked telepresence system with a head mounted display system. The need for a low latency period to achieve high operator performance and comfort during certain tasks is also established. This is evident during trajectory following experiments where the operator is required to track a highly dynamic target.

The telepresence system has been fully evaluated and demonstrated to enhance operator spatial perception via a sensation of visual immersion in the remote environment.
ACKNOWLEDGEMENTS

I would like to thank my family and friends for their encouragement and support throughout this period of study. Their patience and understanding is greatly appreciated.

Thanks must also be made to particular members of the Mechatronic Systems and Robotics Research Group, particularly to Dr. Nongji Chen and Mr. Gary Bonser for undertaking many fruitful and often long technical discussions. My thanks must also be made to Dr. John Pretlove, my supervisor, for his help during this period of study.

Finally, the financial support of the Defence Research Agency and the Department of Mechanical Engineering at the University of Surrey for the funding of this work is gratefully acknowledged.
THE DESIGN, DEVELOPMENT AND EVALUATION OF AN
ACTIVE STEREOSCOPIC TELEPRESENCE SYSTEM

Abstract................................................................................................. i
Acknowledgements........................................................................ ii
Table of Contents............................................................................ iii
List of Figures.................................................................................... vii
List of Tables..................................................................................... ix

CHAPTERS

I. Introduction......................................................................................... 1
  1.1 Introduction................................................................................ 1
  1.2 Achieving Visual Telepresence.............................................. 6
  1.3 Sensor Control........................................................................ 11
  1.4 Summary............................................................................... 12
  1.5 Outline of The Work............................................................ 13
  1.6 Organisation of The Thesis................................................. 16

II. A Review of Telepresence Systems..................................................... 18
  2.1 Introduction........................................................................... 18
  2.2 The Establishment of Head Coupled Systems.................. 19
  2.3 Review of Current Systems and Essential Components...... 25
  2.4 Summary............................................................................ 39

III. Stereo Perception and Display Devices............................................. 42
  3.1 Introduction........................................................................... 42
  3.2 Physiology of the Human Eye............................................. 43
      3.2.1 The Monocular Visual Field.................................... 45
      3.2.2 The Binocular Visual Field.................................... 46
  3.3 Stereoscopic Properties of Human Vision...................... 46
      3.3.1 Stereoscopic Acuity and Stereoscopic Range......... 48
      3.3.2 Human Visual Stereoscopic Space......................... 48
  3.4 Stereoscopic Camera Systems........................................... 50
      3.4.1 Stereoscopic Space - As Perceived by Camera........ 53
  3.5 Stereoscopic Display Devices............................................. 54
      3.5.1 Non-Immersive Separate Image Systems.............. 55
      3.5.2 Non-Immersive Displays with Combined Images..... 55
3.5.2.1 Spatial Overlaying Techniques ...............................56
3.5.2.2 Temporal Sequencing Techniques ..........................57
3.5.3 Immersive Display Systems .................................................59
3.6 Summary ..................................................................................................61

IV. Mechatronic Design Considerations ..........................................................63
4.1 Introduction.............................................................................................63
4.2 Mechanical Design Requirements .......................................................64
  4.2.1 Static Mechanical Design Considerations ...........................64
  4.2.2 Kinematic Design Considerations .........................................65
    4.2.2.1 Range, Velocity and Acceleration of Each Degree of Freedom .................................68
4.3 Vision System Requirements ...............................................................70
  4.3.1 The Camera System .................................................................70
    4.3.1.1 Camera Lens Considerations ........................................73
  4.3.2 The Display Systems................................................................75
4.4. The Head Tracking Sensor System ...................................................77
  4.4.1 Operation of AC Magnetic Tracking System .......................82
4.5 Actuation System Considerations ......................................................85
  4.5.1 Actuator Position Feedback ..................................................86
4.6 Control System Considerations ...........................................................89
4.7 Summary ..................................................................................................93

V. System Design .................................................................................................95
5.1. Introduction.............................................................................................95
5.2 Mechatronic System Design ...............................................................95
  5.2.1 Mechanical System Design ...................................................98
5.3 Control System Design .........................................................................103
5.4 The Controller Architecture ...................................................................105
  5.4.1 Digital to Analogue Converter (DAC) Card .........................106
  5.4.2 Encoder Interface Board ........................................................107
    5.4.2.1 Differential Driver Interface Board .......................109
  5.4.3 Digital Input/Output Board ...................................................110
    5.4.3.1 Interrupts ..................................................................111
  5.4.4 Serial Communications ..........................................................113
5.5 Implementation of the Controller ........................................................115
  5.5.1 Self Calibration ......................................................................118
5.6 The Display System Implementation ................................................118
5.7 Summary ..................................................................................................120
7.7 Summary .................................................................................................. 185

VIII. Conclusions and Future Research ......................................................... 186
  8.1 Introduction ............................................................................................. 186
  8.2 Conclusions ............................................................................................ 187
  8.3 Future Research Directions .................................................................... 189

BIBLIOGRAPHY .......................................................................................... 192

APPENDIX A : Schematic Drawing ............................................................... 197
APPENDIX B : Motor Selection Calculations ............................................... 199
APPENDIX C : Stereoscopic Magnification ..................................................... 207
APPENDIX D : Preliminary System Model ....................................................... 210
APPENDIX E : Formulation of Governing Equations ...................................... 212
APPENDIX F : Modelling Error ..................................................................... 221
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Generalised Schematic of a Teleoperator</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>An Example of Remote Teleoperation</td>
<td>3</td>
</tr>
<tr>
<td>2.1</td>
<td>A Schematic of the Goertz Head Controlled System</td>
<td>24</td>
</tr>
<tr>
<td>2.2</td>
<td>Functional Layout of the Turing Institute Head</td>
<td>29</td>
</tr>
<tr>
<td>2.3</td>
<td>Tachi Head System</td>
<td>32</td>
</tr>
<tr>
<td>2.4</td>
<td>Functional Schematic of NASA System</td>
<td>35</td>
</tr>
<tr>
<td>2.5</td>
<td>BOOM and the NASA System</td>
<td>38</td>
</tr>
<tr>
<td>2.1</td>
<td>A Simplified Schematic of the Human Eye</td>
<td>44</td>
</tr>
<tr>
<td>2.2</td>
<td>The Vieth-Muller Circle</td>
<td>49</td>
</tr>
<tr>
<td>2.3</td>
<td>Two Camera System - Stereoscopic Depth and Magnification</td>
<td>52</td>
</tr>
<tr>
<td>3.1</td>
<td>Simple Model of the Human Neck and Eyes Rotational Degrees of Freedom</td>
<td>66</td>
</tr>
<tr>
<td>3.2</td>
<td>Functional Schematic of the Polhemus System</td>
<td>83</td>
</tr>
<tr>
<td>3.3</td>
<td>Joint Actuator/encoder Configuration</td>
<td>86</td>
</tr>
<tr>
<td>3.4</td>
<td>Incremental Encoder Signals</td>
<td>88</td>
</tr>
<tr>
<td>3.5</td>
<td>Schematic of the Control System</td>
<td>90</td>
</tr>
<tr>
<td>5.1</td>
<td>3D Studio Model of Telepresence Head</td>
<td>99</td>
</tr>
<tr>
<td>5.2</td>
<td>Schematic of Bearing Arrangement for Elevation Axis</td>
<td>101</td>
</tr>
<tr>
<td>5.3</td>
<td>The Actual Telepresence Head</td>
<td>102</td>
</tr>
<tr>
<td>5.4</td>
<td>A Schematic of Control System Configuration</td>
<td>103</td>
</tr>
<tr>
<td>5.5</td>
<td>DAC Bi-polar Transfer Function</td>
<td>107</td>
</tr>
<tr>
<td>5.6</td>
<td>Operation of 82C53 in Rate Generator Mode</td>
<td>111</td>
</tr>
<tr>
<td>5.7</td>
<td>A Functional Schematic of the System Controller</td>
<td>115</td>
</tr>
<tr>
<td>5.8</td>
<td>System Undergoing Head Tracking</td>
<td>117</td>
</tr>
<tr>
<td>5.9</td>
<td>Schematic of Display System Configuration</td>
<td>119</td>
</tr>
<tr>
<td>6.1</td>
<td>Pan Axis Step Response</td>
<td>122</td>
</tr>
<tr>
<td>6.2</td>
<td>Elevation Axis Step Response</td>
<td>123</td>
</tr>
<tr>
<td>6.3</td>
<td>Pan Axis Approximate Constant Velocity Trajectory Tracking</td>
<td>127</td>
</tr>
<tr>
<td>6.4</td>
<td>Pan Axis Dynamic Trajectory Tracking</td>
<td>128</td>
</tr>
<tr>
<td>6.5</td>
<td>Elevation Axis Dynamic Tracking</td>
<td>128</td>
</tr>
<tr>
<td>6.6</td>
<td>Pan Axis Time Shifted (60ms) Trajectory</td>
<td>129</td>
</tr>
<tr>
<td>6.7</td>
<td>Pan Axis Time Shifted (60ms) Trajectory</td>
<td>129</td>
</tr>
<tr>
<td>6.8</td>
<td>Elevation Axis Time Shifted (60ms) Trajectory</td>
<td>130</td>
</tr>
<tr>
<td>6.9</td>
<td>Stereoscopic Magnification</td>
<td>132</td>
</tr>
<tr>
<td>7.1</td>
<td>Teleoperation Work Cell</td>
<td>137</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 4.1 Summary of Mechanical Head Kinematic Requirements .................. 70
Table 4.2 EyeGen3 Specification Summary............................................................. 76
Table 4.3 Summary of Fastrack Performance ..................................................... 84
Table 4.4 Summary of Design Criteria ................................................................. 94
Table 5.1 Summary of Load Inertia, Referred to the Relevant Axis ............... 96
Table 5.2 DC Servo Specifications............................................................................ 98
Table 5.3 Resolution of Each Axis ....................................................................... 108
Table B.1 Vergence Axis Inertial Loading ........................................................... 202
Table B.2 A summary of the RH-5 Motor Specification ................................... 203
Table B.3 Available Encoder Resolutions............................................................. 206
Table C.1 Stereoscopic Magnification as a Function of I and V .................... 209
CHAPTER 1

INTRODUCTION

1.1 Introduction.

For many years it has been man’s desire to have a substitute which can take over those jobs which are viewed to be difficult, dangerous or repetitive. With many of the advances in computing and electronics in recent years these tasks are slowly being performed by automated machinery and robotics. However there still remain many limitations to the current performance of robot systems. Both industrial, and increasingly service robots, are now capable of performing repetitive tasks extremely well, but to perform these tasks autonomously often requires the environment to be either highly structured or constrained. It is apparent that current technology will not allow robots to perform to their full potential when subjected to a highly unstructured environment.

In manufacturing systems, unstructured environments do occur to a certain extent. It is often the case that a particular aspect of the system is known to be unstructured and a sensor solution specific for this task is developed, such as machine vision to locate components on printed circuit boards. Unfortunately this approach cannot be adopted for hostile or hazardous environments as these types of environments are often highly unstructured and sometimes unknown. For this reason an alternative to the fully autonomous type of robotic system is utilised which are known as telerobots or teleoperators.
These systems rely on the input from a human operator and for this reason are often referred to as "man-in-the-loop systems". A critical feature of a teleoperator system is the ability to project man's manipulative capacity into a hazardous or hostile environment, thus providing the operator with a sense of presence in the remote environment. This can be likened to overcoming barriers between the remote environment and operator space as generalised in figure 1.1.

![Figure 1.1 Generalised Schematic of a Teleoperator.](image)

There are three independent components of remote presence (Sheridan 1992) which each contribute to achieving a sense of remote presence, either in a real or virtual environment:

i) Sensory information:

The same sensory information must be provided to the operator as they would receive if they were actually present in the remote environment.
Chapter 1: Introduction.

ii) Control of Sensors:
It must be possible to move the sensing devices around the environment.

iii) Remote Environment Interaction:
The ability to modify the remote environment and manipulate any objects within it.

To achieve the ultimate sense of being present in a remote environment requires the full combination of all these aspects and is essentially the aim of most teleoperator and telepresence systems, although it has generally proved very difficult to convincingly achieve them all simultaneously within one definitive system due to the complexity of the individual systems.

![Image of remote teleoperation setup](image)

**Figure 1.2 An Example of Remote Teleoperation.**
A conceptual example where Sheridan's concepts can be seen in action is shown in figure 1.2. This example system is representative of a teleoperated robot being used in a nuclear cell to manipulate radioactive canisters. The remote robot is controlled by the operator who views the actions of the robot via the remote camera system. The camera system very simply demonstrates Sheridan's first aspect of presence, by presenting the operator with a view of the remote work space, which if the display device was capable of displaying high definition colour stereoscopic images would be very similar to the images that the operator would perceive if present at the remote site. In addition to the visual feedback, the operator gains a further perception of the remote environment through the force reflective controls which generate a representation of the forces being experienced by the robot gripper to the operators hands. This example demonstrates only two channels of sensor communication between the operator and the remote cell, but there are no reasons why this could not be increased to include aural sensing for example. The aim of all the types of sensing is to form a representation of the perception which the operator would be able to gain if they were actually present in the remote environment.

The justification for the requirement of a high degree of sensory presence at a remote operating site is established when the operational tasks are primarily in situations where the tasks are wide ranging, complex and uncertain (Held and Durlach 1992). These typify the conditions which are often found in hazardous or hostile environments such as the nuclear environment demonstrated in the example of figure 1.2. The example also shows that the sensory feedback can take many forms including tactile, force feedback, sound and vision.
Visual feedback or operator visual presence has proved to be the most difficult to convincingly achieve, due to the high sensitivity of the human vision system to effects such as resolution and blurring. However, visual feedback is very important as it forms one of the largest sensory inputs to the operation of the human brain. This has proved so difficult to achieve as essentially, remote presence systems are attempting to artificially transfer or extend the operators sensory system to the remote operating environment. In this thesis it is the formulation of the visual feedback of a remote presence system which is of specific interest.

Sheridan’s second proposed attribute of presence, control of the sensor systems, is also conveyed in the example given in figure 1.2, through the operator being able to control and move the view of the camera system around the work space. This is achieved by manipulation of the second robot manipulator using a separate group of joystick controls to the force reflective controls used to control the manipulator with the gripper. The remote sensors should be able to respond to the demands placed on them by the operator and consequently requires a highly responsive and well controlled system. Furthermore, if the interaction between the sensors can be made intuitive and natural the operator work load is reduced and sensation of presence at the remote site is further enhanced. When there is a combination of visual presence and sensor control in remote environment it is otherwise known as active telepresence.

The third component of presence discussed by Sheridan reflects on the ability of the operator to interact with the remote environment. In this example the interaction with the remote environment takes the form of a robotic manipulator, but this could equally be a remotely operated vehicle or gantry mounted manipulator depending on the application.
A remote vehicle is often used in applications such as bomb disposal, where the manipulator and sensory devices are manoeuvred to the remote site. The design of these vehicles and manipulators and their respective control strategies are very important and are the focus of much research such as that performed by Lee (1993) and Borenstein (Borenstein and Koren 1990). This thesis does not investigate these components of a teleoperator system, but focuses on the generation and control of a stereoscopic sensor system for remote telepresence applications which draws together the first two attributes of presence discussed by Sheridan.

From this discussion it is possible to specify the fundamental features of such a telepresence system. It must consist of a sensor system which can collect visual information from the remote environment and a suitable method of displaying this information to the operator. The display methodology must enable an intuitive and natural interface to be made between the operator and the display device. Therefore an idealised visual telepresence system should replicate the human vision system at the remote location and present this in such a way that the operator would have no basis for distinguishing the real remote environment from the real local environment (Loomis 1992). In order to achieve this requires not only high quality visual feedback but also the correct sensor motor interactions between the operator and the remote system.

1.2 Achieving Visual Telepresence.

The previous section has demonstrated that visual telepresence is an important tool when used as an aid to perform remote operations in environments which are particularly unstructured or hazardous to humans.
This section outlines some of the techniques which can be adopted to capture the images for visual presence.

In some respects the problem of a vision system for telepresence applications is much simpler than for artificial intelligence (AI) applications. As opposed to AI where complex algorithms must be performed in order to interpret the scene, the telepresence system must present a convincing representation of the scene to the human. It is then the human processor, i.e. the brain, which performs computations, scene analysis and interpretation.

When directly viewing any given scene the human uses many cognitive processes which have been built up through many years of experience. In general the human can perform many different assessments of the scene such as (i) estimate the size and apparent distance of objects from the viewing position, (ii) co-ordinate the current view of an object with a previous view as part of a decision making process, (iii) make an assessment of the colour content of the scene, and (iv) build three dimensional interpretations of the scene. This list is by no means exhaustive, but merely highlights some of the processes which are performed when a human views a scene. Therefore in order to achieve a sensation of visual presence in a remote location it would be expected that these natural operations should be able to occur freely. The ability of these operations to take place is highly dependent on the method by which the remote scene is displayed to the operator. Consequently the method of displaying the remote environment also governs the technique by which the scene is sensed or viewed.

There are essentially six different methods by which attempts have been made to generate visual feedback for telepresence applications;
(i) A single conventional black and white television display fed by a single fixed television camera.

(ii) A single conventional colour television display fed by a single fixed colour television camera.

(iii) The use of numerous cameras placed at various positions throughout the workspace.

(iv) Black and white stereo display system fed by a pair of black and white cameras. The separation of the cameras may be equal to or considerably larger than the inter-ocular separation of the eyes of the viewer.

(v) A colour stereo display system fed by a pair of colour cameras. The separation of the cameras may be equal to or considerably larger than the inter-ocular separation of the eyes of the viewer.

(vi) Head mounted stereoscopic display (HMD) fed by a pair of colour cameras.

A single black and white or single colour camera system offers the simplest solution. The advantage of using a colour display is quite obvious in certain applications. For example, if the teleoperator was required to perform an operation on a particular colour wire within a bank of many different wires, then this task would be almost impossible to perform using solely a black and white television system. Although colour systems do offer significant advantages when used in situations where the information provided by colour is of importance, generally it was found that colour systems only offered a very slight increase in performance capability (Freedman 1977). However, the human vision system does recognise the colour content of a scene, and reducing this to only black and white limits the natural understanding of the
scene. Consequently, this also has the effect of limiting the feeling of presence within the remote environment.

Even when utilising a colour camera, the view of the environment is very constrained if the cameras are fixed in position. It also does not offer much flexibility in the range of operations which can be performed. Single cameras and displays also limit the perception of depth that the operator can experience from viewing the scene. Limited depth perception is achieved via depth cues such as shadowing or linear perspective (Reinhart, Beaton and Snyder 1990). However, human vision is based on viewing an environment from two sources - the eyes. The brain then fuses the information from both eyes to form a three dimensional model of the space in question. Here lies an obvious limitation with the use of single camera systems, the remote environment is only sensed from a single position and hence the display can only be monoscopic.

The use of numerous single cameras placed throughout the workspace is somewhat artificial and does not lead to a natural understanding of the remote environment. Often, in order to localise an area of interest, it is necessary to mentally integrate the views from numerous camera viewpoints and infer the position of interest. It is often found difficult to achieve this without much experience in using such a system or if the operator has little spatial understanding. This technique does not offer any true ability to generate telepresence although with sufficient skill, it is possible to infer three dimensional spatial orientation.

Stereo display systems simulate a sense of depth by presenting two disparate images, one to each eye. That is two images which have been formed by sensors looking at the same points in a scene from slightly
different positions. There are many techniques of presenting stereo images generated from two separate sources which will subsequently be discussed in more detail.

Many different experiments have been undertaken (Miller and Mitchell 1990, Cole and Parker 1989) to validate the use of stereoscopic based systems. It is evident that the usage of stereoscopic display systems improves the operator performance, particularly if the operator is not experienced in the use of monoscopic systems. Also three dimensional display systems offer significant advantages when utilised in either constantly changing or poorly illuminated environments, as is experienced in many sub-sea applications. In situations where there is less than optimum lighting conditions, the added informal cues experienced from stereopsis dramatically improves the overall system performance (Drascic, Milgram and Grodski 1989).

Head mounted display (HMD) systems are an alternate type of stereoscopic viewing system which are worn over the eyes of the user and generally give the wearer the impression that they are immersed within the environment displayed in the images. Unfortunately, HMDs are not exempt from current liquid crystal (LCD or cathode ray tube (CRT) technological limitations and subsequently offer limited resolution and image quality. HMD systems are available in many different formats which will subsequently be discussed in more detail.

For improved operator performance the visual system of the telepresence system should enable a stereoscopic view of the remote environment to be formed. An ideal technique to present the stereoscopic images to the operator is to use a HMD. This allows the aim of visual immersion in the
remote environment to be achieved and forms one crucial feature of the
design of a stereoscopic telepresence system.

1.3 Sensor Control.

The presentation of images to an operator via a HMD may generate a feeling
of being visually present in a remote environment. However, as soon as the
operator changes their head position if these images do not change
accordingly the feeling of presence will be lost. This is due to the lack of
interaction between the operator and the remote camera system. The
interaction between the operator’s head motion and the camera system
motion is a further crucial feature of a high performance telepresence
system (Sheridan 1992).

In order to convince the operator that they are immersed in the remote
environment the visual image presented to the operator must be in relation
to the user’s head position. It must be possible for the operator to sweep the
direction of gaze by rotating the head and the visual scene which is projected
onto the human retina to change appropriately. This can be achieved by two
distinct approaches (i) by using a hand controlled joystick to control the
motion of the cameras or (ii) by coupling the operator’s head motion and the
remote camera system. However, as explained in section 1.2, according to
Sheridan (Sheridan 1992) an ideal telepresence system should not only offer
the ability to control the sensors in the remote environment but also allow
this to be achieved via a natural interface. Consequently the ideal solution to
this control problem is to control the orientation of the remote camera
system via the operator’s head motion. Hence, the remote camera system
can view the remote scene and present it to the operator via a head mounted
display system (HMD).
Through tracking the operator head position, the remote camera system can be controlled in such a way that the head motion is replicated. The effect of this correlation between the operator and the remote mechanical system is to further enhance the sensation of telepresence. However it is significantly degraded if the system cannot rapidly respond to the head motion trajectories demanded by the operator. This demands that the telepresence system has a very short time delay between the onset of the operator head motion and the camera motion, which is otherwise known as the latency period. One of the aims of this work is to develop a telepresence system which has a far reduced latency period compared to other current systems. The latency period and its effects on operator performance are investigated to justify the necessity for reduced latency period.

1.4 Summary.

The possibility of distancing an operator from a hazardous or hostile environment is not a new technique. However, previous systems do not convey a sense of real remote presence in the environment, although they would have certainly been an aid to perform the task for which they were being used. To exhibit a sense of remote presence according to Sheridan (Sheridan 1992) necessitates three characteristics of a system (i) a remote sensor system, (ii) enable manipulation of the sensors and (iii) allow interaction with the remote environment. Telepresence can be achieved by a combination of the first two characteristics, specifically being able to manipulate a vision system in a remote environment. This sensation of presence, or telepresence, is particularly beneficial when the operating space is rapidly changing and the tasks are wide ranging and complex, as in bomb disposal for example. Telepresence is also well suited to assisting operation in
very unstructured and unpredictable environments where currently full autonomy is still impossible to achieve.

In order to convincingly achieve telepresence, there are two main features a system must possess. Firstly, the technique of displaying the remote scene to the user must enable an intuitive understanding to take place. The most natural form of this is to utilise three dimensional display technology. Secondly, it is very important to have the correct motor interactions between the human operator and the remote system. Again this must be achieved in a natural and intuitive manner if convincingly high performance telepresence is to be demonstrated. These features outline the theoretical requirements described by Sheridan (1992) namely that a remote presence system must provide sufficient sensory information whilst allowing the sensors to be controlled according to the demands of the operator.

1.5 Outline of The Work.

The aim of this work is to design, develop and evaluate a stereoscopic telepresence system. In doing so, the work addresses a number of issues in order to overcome some of the limitations of other telepresence systems. It concentrates on: a) the formulation of a design specification for a high performance telepresence system including an evaluation of the system dynamic requirements, b) the mechatronic design of the system, c) the design and development of a PC/AT based open architecture controller, d) evaluation of the telepresence system performance both in mechatronic and operator oriented terms.

The main limitation with current telepresence systems is their poor dynamic performance and inherent latency which is often in excess of 170ms.
Excessive latency periods can cause the operator to become disorientated due to the camera system not immediately realising the change in operator viewpoint. This effect is highly detrimental to the sensation of remote presence and should be minimised. The effect of prolonged latency periods on the performance of a task by an operator is demonstrated in this thesis.

Whilst demonstrating a much reduced latency period, the system must also be capable of replicating the dynamic motions of the human head, such that any user generated motions can be simulated. However, for the purpose of this initial investigation the motion is to be limited to the elevation and panning motion. A two degree of freedom neck motion offers a more simple design and the cost of a third degree of freedom could not be justified for this initial investigation.

The performance of the telepresence system is to be such that the dynamic characteristics of the mechanical system are similar to those of a human operator. This will ensure a high degree of operator immersion. Through the development of a dedicated open architecture controller incorporating real-time techniques, a suitable solution was developed.

The open architecture controller is very easily configured to accept inputs from a variety of sensors, such as joysticks, keyboards or tracking sensors. Many of the current systems are only able to accept a single form of sensory input. Enabling the possible connection of various input devices would allow the performance of different system configurations to be evaluated. However, for the purpose of this work the sensor input was confined to a Polhemus Fastrack sensor, a digital joystick and a keyboard entry system.
A further limitation of other systems is that they exhibit a prohibitively large space envelope. It was therefore a requirement to minimise the space envelope of the system such that it could be mounted on to a small laboratory mobile vehicle. This produces constraints, both in the physical size and mass of the system, demanding that the system weigh less than 10Kg and occupy a cylindrical space envelope with dimensions not exceeding 300mm height X 300mm diameter.

The implementation of this system allows the advantages of using a stereoscopic head coupled system to be demonstrated. The system is to be used to perform the role of a visual interface for a telerobotic manipulator. The enhanced operator spatial perception which is achieved through using head coupled stereoscopic camera systems over conventional monoscopic systems will be demonstrated. This can be quantified by a reduced task completion time and improved accuracy.

Experiments are performed to clearly demonstrate the advantages of using a head coupled system over a joystick controlled system. This is shown by improved operator performance and is demonstrated by a series of target tracking experiments. A comparison is also made between immersive and non-immersive operation of the telepresence system.

Further experiments investigate the effect of increasing the time lag between the motion of the operator's head and the remote telepresence system, known as latency. The magnitude of this delay is particularly important from an operator perspective. If the delay is too large then during rapid operator head motion the telepresence system becomes difficult to use and sensations of motion sickness may be experienced. Consequently this design has been aimed at minimising these effects through a reduction of...
the latency period. This is achieved by an integrated mechatronic design approach.

These experiments aim to effectively demonstrate and validate the system design. They will also demonstrate the suitability for use as an integral component of a teleoperator cell or an immersive remote inspection system.

1.6 Organisation of The Thesis.

This thesis is organised as follows. In Chapter 2, a review of the development of telepresence systems is presented. This outlines the historical development of telepresence systems and concludes with a review of the current 'state-of-the-art' systems and essential components.

In Chapter 3, a review of the physiological operation of the human eye, in both stereoscopic and monoscopic operation is presented. This chapter will also demonstrate the mathematics of stereoscopic camera systems utilising two cameras and review the various methods which can be used for the display of stereoscopic images.

In Chapter 4, the design considerations for the telepresence system are presented. These are used to evaluate the requirements of all the various sub-systems from which the telepresence system is to be comprised. Within this, the mechatronic aspects of the design and the kinematic requirements of a high performance telepresence system are also outlined.

In Chapter 5, the actual design and implementation of the system is presented, with a particular emphasis on the design of the real-time controller.
Chapter 6 evaluates the mechatronic oriented performance of the system. This includes the dynamic response and the optical characteristics. Experiments are also performed to identify the overall system latency period.

In Chapter 7, investigations are presented which characterise the task oriented performance of the telepresence system. This is achieved by a series of experiments using the system as a visual interface of a telerobotic cell. A further target tracking task is used to evaluate and compare the operator performance achieved whilst using a range of demand input and display devices. The effects of increased latency periods on the operator performance are also investigated using the same target tracking task.

The conclusions which can be drawn from the research and suggestions for future research are presented in Chapter 8.
CHAPTER 2

A REVIEW OF TELEPRESENCE SYSTEMS

2.1 Introduction.

Telepresence systems are required to convey a sense of presence within a remote environment. This can be achieved by supplying the system operator with sufficient quantity and quality of sensory information or feedback to approximate actual presence at the remote task site. It is particularly advantageous to generate a sense of remote presence when the tasks to be performed are very varied or alternatively if the environment in which these are to occur is unstructured or unknown. Unstructured and unknown environments frequently occur when the environment is hazardous or hostile, such as that experienced in sub-sea or nuclear decommissioning applications. In an attempt to satisfy the requirement of telepresence a number of systems have previously been developed.

Each of these systems has to incorporate many considerations into their design due to high performance telepresence placing stringent demands on the system. Many requirements if not achieved satisfactorily are detrimental to the sensation of remote presence. Head coupled systems with large time delays between the operator motion and the mechanical system motion prevent the sensation of presence being achieved and can induce a sense of disorientation to the operator. Also low resolution cameras and display devices do not allow an adequate perception of the remote scene to be formed.
This chapter reviews the historical establishment of telepresence systems and reviews the research and development of recent systems. It focuses on stereo vision platforms which are or could be used for telepresence applications. Although some of the systems reviewed may not be regarded in their entirety as complete telepresence systems, their design exhibits features required by such systems. In some cases these systems are used in applications where a form of telepresence is attempted, such as using a stereoscopic display to observe a manipulator performing a nuclear decommissioning task. In other cases the telepresence camera systems are similar in design to those used for undertaking computer vision and AI research. A detailed review of AI oriented systems is not presented, however, an in depth review of this type of stereo vision research platform is presented in a thesis by Pretlove (Pretlove 1993).

2.2 The Establishment of Head Coupled Systems.

The first drive towards teleoperators was made in the late 1940s under the impetus of the United States Atomic energy programme. This application was the first to necessitate the projection of man’s manipulative capability into a hazardous environment. The first manipulators were unilateral and tended to be used in a direct viewing capacity, where the operator could directly view the actions and spatial relationship of the manipulator. This type of teleoperator allows force and position to be transmitted from the operator to the manipulator only.

There are several difficulties with the use of direct viewing systems for hazardous environments, concerning both environmental and many human operator related problems. The physical limits to direct viewing are highly environment dependent: for example submersibles for reaching depths in
excess of two thousand metres can only tolerate small hull penetrations (Johnsen 1971). Also in 'hot' nuclear environments it is preferred to limit the occurrence and size of wall penetrations, consequently the windows are frequently stepped or flared such that they open up towards the inside of the nuclear cells. Thus limiting the view which the operator has of the remote environment.

During 1964 the American Airforce Medical Research Laboratories investigated the human factors aspects in relation to direct viewing teleoperation and established that the distance between the operator's eyes and the work piece should not exceed three and a half metres due to the reduction in human depth perception and visual resolution at any greater distance (Kama 1964). More importantly, these factors have the effect of increasing completion time and the failure rate of certain tasks. The method of overcoming this problem was to allow the operators to utilise binoculars or telescopes for operations at greater depths and was often implemented in many of the larger nuclear cells.

During the mid 1960s, the requirement for more sensory feedback from the manipulator was realised with the development of bilateral master-slave manipulators (Goertz et al. 1966) at the Argonne National Laboratory (ANL). Bilateral teleoperators allow both force and motion to be transmitted from the operator controls to the actuators and vice versa. In it's simplest form force feedback is generated in a bilateral teleoperator by back driving the master driving actuator to create a position offset. Even in their earliest form bilateral manipulators exhibited a significant decrease in task completion time when compared to unilateral manipulators (Martin and Hamel 1984). This is rather dependent on the type of the task
which is being undertaken and it would be expected that for very simple tasks that force reflection would not exhibit such an advantage.

Through the generation of force and position feedback, bilateral manipulators were slowly beginning to improve the sensation of operator remote presence. Manipulators continued to be further developed, both unilateral and bilateral, and are still under development today, with particular emphasis on increased performance and simulation. Nowadays bilateral manipulators are increasingly utilised in many application areas such as space and the nuclear industry.

It was soon realised that use of telerobotics with the facility of directly viewing the manipulator actions was somewhat constraining. It was often found difficult to perform operations when they were to be done at a considerable distance from the windows, or when it is necessary to view the work from a direction not permitted by the windows. Indeed in some cases it was just not physically possible to perform direct viewing of a manipulator, such as deep within a nuclear pile or at great ocean depths.

As a consequence, during the 1960s, experiments were performed using conventional 2D black and white television as the visual sensing medium. The superior and practical utility of television systems when used to complement teleoperation was soon realised. Television systems are inherently more portable and offer the ability to allow control of a manipulator at distances far exceeding the range of direct vision. By including a number of cameras it was possible to generate views from a number of positions, which may be widely separated in space. However there were also some fundamental problems associated with the use of conventional 2D black and white television systems at this time. The
systems tended to be of poor resolution, required a very large communications bandwidth and offered poor operator depth perception.

The only shortcoming which was directly approached at that time was the problem of depth perception. The Argonne National Laboratory (ANL) experimented with 3D black and white viewing systems. The systems were not very successful due to the poor performance of the equipment and subsequently the whole principle of television viewing was frowned upon. Another contributing factor for the disapproval of 3D systems was the fact that they required rather large cumbersome camera arrangements and also added extra burden onto the operators.

The operators were required to control two camera systems in addition to the manipulator which was not an intuitive or natural interface. Just as the manipulator is connected, indirectly, to the operators hands as a natural interface, it is preferable for the operator to naturally control the camera system. In doing so the operator would be relieved of the additional task of ensuring optimum camera positioning. The most intuitive interface was to link the head position of the human operator with that of the camera system and was first achieved by Philco Corporation (Comeau and Bryan 1961), although this only operated as a monoscopic system. This was then referred to as a head slaved or head coupled system.

The system developed by Comeau consisted of a close circuit television surveillance system which was used with a miniature CRT television display. The image from the display is presented to the operator via a mirror which is placed in front of the eyes.
By using a magnetic head position sensing system the operator's head position was used to control the position of camera unit. The camera unit featured a single vidicon television camera mounted on a pan and elevation unit and consequently only a monoscopic image could be formed. The positioning sensing system was based on a pair of coils mounted on the operator's head and a set of Helmholtz coils placed around the head. The coils mounted on the user's head sensed the phase of three magnetic fields, thus enabling the pan and elevation of the head to be calculated. The gaze direction of the camera was then modified to match the change in position of the operator’s head.

The success and acceptance of this system was quite limited. It suffered from problems with the resolution of the magnetic sensor and the display device. However, it was the initial step in an attempt to create telepresence. The authors also recognised many future applications for telepresence such as sub-sea, space exploration, radioactive contaminated areas and military surveillance.

In 1965, the Argonne National laboratory (ANL) developed an alternative head controlled viewing system (Goertz et al. 1965), for use in the atomic energy programme, shown in figure 2.1. Unlike the Philco Corporation system, the display of the images was not presented directly to the operator's eyes by the use of mirrors, but instead utilised a conventional television monitor. The television system formed part of a commercially available close-circuit system using a single vidicon camera.
This experimental system allows the operator to move their head in any direction, but only the pan and elevation motions control the position of the camera system. The operator’s head position controls the position of the viewing monitor to which the camera is slaved in a 1:1 relationship. The monitor is supported on a boom which moves in a spherical manner centred around the operator which maintains the face of the monitor pointing at the operator’s eyes at all times. The motion of the monitor and camera are
driven by conventional servo motors which are capable of achieving a maximum rotational velocity of 0.5 rad/sec (28.6 °/sec).

To control the position of the monitor and camera the operator wears a counterbalanced head piece, with which the operator is supposed to feel little restraining force. Synchros attached to this head piece generate a demand signal which is subtracted from the synchros attached to the monitor. The resultant voltage controls the position of the monitor. The camera is then controlled by the comparison of the monitor and camera position signals. The servos were set up initially to accurately follow the operator head position, however during testing it was found that inadvertent head motions caused a surprising amount of motion in the pan and elevation directions. To eliminate this effect, a dead zone ranging between 7 to 12° of the operator’s head motion was introduced. Also during head motions of greater than 0.05 rad/sec (2.8 °/sec) there was some blurring of the monitor display due to the time constant of the Vidicon camera and interlacing of the scan lines on the monitor. Despite these shortcomings, the system developed by Goertz set the scene for active telepresence and its future development.

### 2.3 Review of Current Systems and Essential Components.

Towards the late 1970s head coupled telepresence systems were still under development, particularly in an attempt to utilise stereo viewing techniques. By this time many monocular head coupled systems were in use (Vertut and Charles 1977). Vertut, of the French Navy, had developed a cable controlled deep submergence teleoperator system designed for performing remote observation, investigation and intervention from a surface ship.
The design of the submersible tried to achieve a symbiotic relation to its operator in head sensory, arm dexterity and body agility aspects. The system incorporated a six degree of freedom platform on the front of the vehicle bearing a single miniature television camera. The degrees of freedom of the platform include three rotational and three coupled translation motions and are said to exhibit motions similar to the head and neck. This is achieved through a design which could be likened to a Stuart Platform.

Head coupling for this system is achieved through a mechanical linkage attached to a helmet. The image produced by a miniature television screen is directed towards the operator's eyes using optics. This enables the rather bulky display to be mounted on top of the helmet. The paper offers no information on the increase in performance and also does not present any information on the resolution or field of view which was used.

Amos (Amos and Wang 1978) reports on a design, which at the time was very unique, utilising a head mounted display and a head position sensor. The main development of this system was the utilisation of a stereo television system in addition to head coupling, using a single three dimensional camera. The head coupling was to be achieved by mounting the camera on a gimballed platform and used a non-contact sensor to track the operator's head movement. The stereo television images are presented to the eyes of the operator through two small head-mounted cathode ray tubes. It is thought that CRTs would have been of very low resolution, most likely black and white and definitely extremely expensive display devices.
This paper does seem to be only suggesting an idealised system as the fundamental details of many of the components are not presented, such as the functionality of the non-contact head position sensor. However, the author does highlight that the individual components which are required to build the system were within the reaches of the technology of that time.

In the same year as the theoretical system proposed by Amos, Berry (Berry and Rice 1978) proposed a new three dimensional television system for use in remote servicing. This paper was not so much interested in the generation of a head coupled system, but was more concerned with the technique of generating stereoscopic images using conventional television cameras and monitors. The system used two standard television cameras with fixed viewing positions. The images of which were viewed on a standard monitor. The stereoscopic effect was generated by multiplexing the images formed by the two cameras and then viewing the monitor with polarising shutter glasses. A more in-depth discussion of stereoscopic display techniques is presented in Chapter 3.

In 1988 an alternative method of capturing three dimensional images was proposed, (Robinson and Shuttleworth 1988) using a single miniature stereo camera. This consisted of two imaging chips as opposed to a single one in most cameras, one for each left and right channels, which are controlled by a single set of drive electronics. The technique of displaying the images was the same as the technique adopted by Berry (Berry and Rice 1978), which involves the inter-lacing of the left and right images onto the television screen which is then viewed through polarising glasses.

The Robinson stereo camera has been designed in such a way that all the optical functions were controlled by a small microcomputer. The functions
which were controlled were (a) focal length of the lenses, (b) the camera separation and (c) camera convergence angle. These parameters are elements which control the integrity of the stereoscopic image which was presented to the operator.

The camera convergence and separation functions were controlled using stepper motors and the lens parameter had its own DC servo motor and feedback potentiometer. The microprocessor produced the relevant demand position, but was only capable of performing a single motion control operation at a time. The microprocessor prevented the operator from attempting to set the camera parameters to provide images which were difficult to view in stereo. The system was initially configured by programming the average tolerance of stereoscopic fusion limit for each user into memory. Consequently the operator would be notified if this limit was being exceeded and the camera parameters changed accordingly.

The paper provides no information on the resolution of the various control loops nor does it detail whether the system could be actively controlled by the operator. On this basis it would be credible to assume that the system was not head slaved in any respect and was solely an experimental three dimensional viewing system.

The UK Atomic Energy Authority (UKAEA) Harwell developed a three dimensional television system for remote handling operations (Dumbreck, Abel and Murphy 1990). The main thrust of this work was to develop a three dimensional camera. The camera includes a mechanism which directly links the convergence angle with the depth of focus. This link, although initially mechanical, was modified to enable direct software control. The position feedback of the focusing mechanism is used to drive
the convergence, this ensured that the plane in space onto which the cameras are converged was in focus. Dumbreck states that the camera can converge and focus down to a near point of 300mm, but there is no statement of the available resolution. The camera uses two re-packaged single-sensor CCD cameras with a range of lenses which makes a field of view angle from 8 to 30° possible.

The operator is only able to control the vergence angle of the CCDs. However, it was found during testing that the whole workspace could not be viewed directly and so warranted additional flexibility offered by a two degrees of freedom pan and elevation unit. The camera images are viewed stereoscopically via two display monitors. The monitors are set at right angles and then viewed through a beam splitting mirror. Polarising filters on the monitors and worn by the viewers separate the images. A limitation with using this type of display is that the stereoscopic effect is easily lost if the viewers head does not remain reasonably fixed in front of the viewing mirror.

![Diagram of Turing Institute Head](image)

Figure 2.2 Functional Layout of the Turing Institute Head.
Many stereo head systems have been developed with applications other than telepresence in mind. A good example of which is the anthropomorphic head which was developed by the Turing Institute (Undebekken 1991). The head features six motorised degrees of freedom with azimuth and elevation control for both the neck and each of the eyes, shown in the schematic of figure 2.2. The principle motivation behind this design was for it to exhibit the degrees of freedom and physical dimensions equivalent to that of the human head. The overall goal was to develop a series of sensory reflexes for the head with the speed and operation which is found in biological systems.

A pair of miniature remote-head colour CCD cameras were used as the image collection devices, and due to this head not being used for telepresence applications there were no external display devices utilised as such. Each degree of freedom is actuated by a stepper motor, the convergence motors allow the cameras to be rotated at a speed of 60° per second with an accuracy of 0.03° per step and the elevation axis can supply a speed of 90° per second with an accuracy of 7.5° per step. The control system for the head employs a PC, within which the various interface electronics are enclosed. The PC communicates to a host Sun workstation via a RS-232 link. This allows the mechatronic system to become independent of the algorithms which perform the reflex behaviours.

Although anthropomorphic in design the Turing head offers many features which would not be required in a telepresence system. It is possible to independently elevate each camera about its own axis which would do nothing apart from disorientate an operator. Also the use of stepper motors has a detrimental effect on the accuracy of the position of the camera system. Stepper motors also limit the speed of the pan and elevation control and would therefore be far too slow for a head coupled application where
the human head can pan and elevate at speeds of greater than 400° and 150° per second respectively (Emsley 1953).

A visual telepresence system was developed as a subsystem of a tele-existence master slave system at MITI, Japan (Tachi, Hirohiko and Maeda 1990a). The tele-existence system is anthropomorphic in form and comprises of a seven degree of freedom manipulator arm and a three degree of freedom neck onto which the stereo camera system is mounted, figure 2.3.

The structure of the robot closely matches the structural dimensions of the human operator. All the three degrees of freedom on the neck are actuated via DC servo motors which utilise harmonic drive gearboxes and incremental optical encoders. The ranges of the neck motion are ±150° pan, ±45° roll and +90°,-30° elevation. The speed of the joints, pan 360°/sec, elevation 240°/sec and roll 240°/sec, are set to match the moderate speed of the human neck motions. All three axes of the neck motions occur around a single position which is located 245mm to the side and 50mm above the shoulder of the robot, which is of similar dimensions to that of a human.

The vision system of the stereo head consists of two colour CCD remote head type cameras. Each camera has 420,000 pixels and uses lenses of focal length 15mm which exhibit a fixed aperture opening of F1.6. The two cameras are set to be aligned parallel, looking in a forward direction with an inter-ocular distance of 65mm. In addition there are microphones mounted on the head system which provide auditory feedback to the operator.
Three methods of head slaving control are proposed by Tachi (Tachi, Hirohiko and Maeda 1989, 90b, 91) (i) A mechanical linkage system with an integral stereoscopic display, (ii) An impedance controlled system again with an integral stereoscopic display and (iii) A head mounted display utilising an electromagnetic sensor.

The mechanical linkage system attempts to minimise the gravitational force from the mechanism to which the operator is subjected through a counter balancing system. The linkage has optical encoders located at the joints and hence by the use of pulse counters the relative position can be generated. This signal is passed to the control computer which forms a closed loop control system with the DC servo and encoder. The visual display device in this system is formed from two miniature CRTs, one for each eye which are mounted on the mechanical linkage in the form of a helmet. Using mechanical linkages as the input device is relatively straight
forward, but the major disadvantage is that the linkage arrangement is very large.

The second coupling mechanism proposed was an impedance controlled system. The system has two degrees of freedom and has an active power assistance mechanism and its impedance can be controlled by an internal feedback loop. Direct drive motors are used in the mechanism and a dedicated computer controls the impedance such that the human operator feels only a low inertia compared to the actual physical inertia of the system. The idea behind this method of coupling was to give the sensation that the system was actually passive and very much lighter in weight. The visual display device used in this method was constructed from two 3 inch colour Liquid Crystal Displays (LCD), the resolution of which was not specified.

The third method of operating the system, in head slaved form, was to utilise a Head Mounted Display (HMD) and an electromagnetic sensor. The HMD was constructed from two 4 inch colour LCD displays, with resolutions of 320 horizontal by 220 vertical pixels. The distinct advantage of using the HMD is that the operator is not tied to a specific location as is demanded by the mechanical system. Unfortunately, at this time, HMDs were considerably large and in this case weighed 1.7Kg. Since there can be no gravitational or inertial compensation with this system it is paramount that the HMD is of light weight construction as all the mass must be carried by the user.

The HMD position in space is tracked using an electromagnetic sensor. This consists of a 3-axis field source and a similarly constructed 3-axis field sensor. The sensor is capable of tracking all six degrees of freedom of
human head motion in real-time. A more detailed explanation of the operation of this type of sensor is given in Chapter 4. The sensor can measure the required position with a position accuracy of ±2.5mm and an angular accuracy of ±0.5° in a ±1.5m cubic measurement field.

As part of the Virtual Environment Workstation project (VIEW) at NASA's Ames Research Centre a remotely controlled stereoscopic camera system was developed as a tool for performing telepresence research (Bolas and Fisher 1990). It was used to evaluate configurations of head-coupled visual systems which are associated with space station telerobotics and remote manipulator arms. The system consists of two cameras mounted on a computer controlled platform which provides pan, elevation and roll control in co-ordination with the head position generated by the user, a functional schematic is shown in figure 2.4.

The system uses two black and white Sony CCD video cameras with a resolution of 384 by 491 pixels. The cameras are very light weight with a mass of around 155g and can accept an external synchronisation signal to ensure correct stereoscopic image presentation. A novel feature of the camera and display system was in the use of very wide field of view lenses (Howlett 1992a). The viewing system utilised optics which provided a 120° field of view to each eye with a 90° binocular overlap. However, these optics introduced distortions to the images such as pin cushioning and aberrations. To overcome these distortions a set of lenses were used on the cameras which would introduce equal and opposite distortions. Achieving such a wide field of view, whilst minimising the distortion of the images, is a definite advantage in a telepresence system especially as the field of view tends towards that of the human. However, due to the low the resolution of the imaging and display system the image quality will appear to be low.
The cameras are arranged with their optical axes parallel and the interocular separation is manually configurable between 38 and 200mm, but it is not possible to modify the convergence angle of the cameras.

The mechanical configuration of the head incorporates three degrees of freedom pan, elevation and roll of the camera system. The camera platform can rotate beyond 360° in all three axes of motion without any limits imposed by the mechanical structure. However this seems to be somewhat excessive, as the human head motions are far more constrained.

![Functional Schematic of NASA System](image)

Figure 2.4. Functional Schematic of NASA System.

All the three degrees of freedom are driven via DC permanent magnet motors with 1000 line optical shaft encoders. The speed reduction between the motors and the load is achieved by using timing belts. Whilst the use of timing belts enables the motors to be located off the axis of rotation there are typical design concerns which must be accounted for, such as backlash,
slip and stretching of the belts. Bolas states that through using fibre reinforced belts with curvilinear teeth these effects should be minimised.

Timing belts allow the elevation and roll axes actuators to be located off the rotational axes and by using a differential gear cluster, the elevation and roll axes share a symmetric drive arrangement. The cable-ways of the system were designed to run through the centre of rotation of each axis such that interference of the cables with the rotational motion of the cameras were minimised.

The motors which drive the platform are controlled by a set of commercially available motion control cards which can perform standard proportional, integral and differential control (PID). Each of the three control cards are synchronised in order to allow for co-ordinated motion between all three axes. The control cards are operated as point to point controllers with the position feedback generated by the optical encoders. The cards are programmed to accept the orientation co-ordinates from the host computer as ASCII strings which are sent over a serial data link. The motion control parameters are calculated as a background process which takes 12 ms to complete, introducing a processing delay. An additional delay of one sample period (66ms) is introduced by the 'on-the-fly' control scheme which is used, requiring the motion control cards to wait and see where they are to move the load to, before setting off to go there.

The motion control cards are connected to the host computer by a RS232 serial data link. It takes 22 ms for a block of orientation co-ordinates to be transferred between the host and the cards, which accounts for a further time delay in the system. The overall latency of the system is noted to be 260 ms when the electromagnetic tracking system and head mounted
display are used and 140ms for the case when the mechanical tracking system is used.

Due to the configuration of the actuator system there is an inherent singularity which occurs when the cameras are looking either directly up or down. This is especially noticeable when the orientations are specified using Euler angles. When the roll axis is in direct line with the pan axis there are an infinite number of ways to specify the orientation in terms of Euler angles. Bolas overcame this difficulty through the use of quaternions to generate the rotational co-ordinates. This was only required when the head tracking device did not exhibit a direct mapping to the mechanical system and two different devices were used to perform head tracking.

The first method of performing head tracking was through the use of an electromagnetic sensor which was located on the operators HMD. The tracking device is connected to the host computer (HP9000) which performs the quaternion computations and data formatting. The host computer is very limited in input/output performance and the co-ordinate throughput of the machine is limited to 15Hz. Nowadays it is possible to program the electromagnetic sensing device to generate absolute orientation co-ordinates, thus reducing the necessity of performing co-ordinate transformations.

The second input device used with the system was a counterbalanced CRT-based stereoscopic viewer (CCSV) (McDowall et al. 1990) which was also developed by NASA Ames, which subsequently became known as the BOOM. Figure 2.5 shows the BOOM and the NASA telepresence system. The CCSV is a mechanical linkage structure, from which the orientation co-ordinates are formed by interpreting the joint positions. Due to the CCSV's
kinematic structure the orientation of physically correspondent joints in the linkage can be directly mapped to the camera system co-ordinates. This input device also enables the singularity to be avoided as the physical joints can only be at one place at a time. This also reduces the number of calculations which the host computer must perform to only a few linear transformations before passing the information to the motion control cards, consequently the co-ordinate throughput rate is increased to 44Hz.

Figure 2.5 BOOM and the NASA System.

The author reports on some of the performance aspects of the system whilst using the electromagnetic sensor and the mechanical CCSV systems as the tracking devices. Two main aspects of system performance were reported, the system static overshoot and the total system tracking error. The static overshoot of the system is determined by comparing the final location of the motorised platform with the motion tracker data recorded during the motion. This showed an error of approximately 1° in 180° for the electromagnetic sensor and approximately 8° in 180° for the mechanical system. The electromagnetic sensor error is thought to be inherent within
the sensing device and does fall within the manufacturers specifications. The total system following error was found to be 260 and 140 milliseconds for the HMD and the CCSV system respectively. The system following error was evaluated through moving the tracking device through a 180° slew at a speed of 140°/sec.

A system recently made commercially available was developed by British Nuclear Fuels (Montgomerie 1994), as an aid for the decommissioning of nuclear plants. The system consists of two cameras which are mounted within a machined aluminium housing and a stereo viewing system. The camera unit does not enable any change in viewpoint to be accomplished by the operator, unless the unit is mounted on some other device, such as a manipulator arm. It seems that the main bulk of the work has focused on the development of the stereo display system which comprises of a monitor with an LCD shutter acting as a polarisor. This system is very similar to other commercially available stereoscopic displays of this type and doesn’t offer any additional functionality.

2.4 Summary.

This chapter described the historical development of telepresence systems from the first developments of head coupled systems by the Philco Corporation (Comeau and Bryan 1961) and the early work performed at ANL (Goertz et al. 1965). Following this a number of more recent telepresence and three dimensional viewing systems were reviewed, some in great detail. This represents the development of telepresence systems from their initial conception to the more modern 'state-of-the-art' systems available today.
Essentially there have been two main thrusts towards the development of remote presence systems; the nuclear industry and space exploration. Both of these environments offer similar challenges for robotic systems due to the distances which are often apparent between the operators and the remote task. In some cases, such as nuclear decommissioning, it is the unstructured and unpredictability of the environment which offers challenges. It is from these areas that the most prominent telepresence systems have emerged, namely the NASA Ames system (Bolas and Fisher 1990) and the system developed by MITI (Tachi, Hirohiko and Maeda 1989). Each of these systems exhibit differences in their configuration and performance.

Both of the systems were designed to match the performance of the rotational kinematics and dynamics of the human head motion. The systems offer three degrees of freedom incorporating pan, elevation and roll motions and the MITI system appears to achieve this in a much smaller space envelope through the use of direct drive motors. The NASA system appears to be rather large, but does enable full 360° rotations on all axes to occur, however this is not particularly an advantage due to an operator's head motion being far more constrained.

It does seem that the inclusion of the roll neck motion complicates the mechanical structure of the stereo platform and also in the case of the NASA system forms singularities during control. The inclusion of a roll motion in the mechanical head allows all the rotational degrees of freedom of the human head to be accounted for, thus all of the operators' rotational head motion will be replicated. During roll motion, the visual content of the scene does not change, it merely modifies the orientation of the scene. In many tasks the utilisation of the roll motion by an operator
tends to be quite minimal, such that for an initial investigative study of telepresence, it would be acceptable not to include the roll axis.

Most of the systems which have been reviewed in this chapter have used commercially available lens systems and it is only the NASA system which used a novel lens configuration.

Very few of the papers reviewed discussed the overall system latency, that is the time taken between the operator initiating a demand for a new orientation or position and the mechanical system beginning to move towards this position. The latency times which are quoted for the NASA head seem to be very high. A latency time in excess of 200ms is unacceptable for a telepresence system (Stone 1991). This must be due to a poorly designed controller and a low performance head tracking system.

All these considerations allow suggestions for an ideal telepresence system to be put forward. An ideal telepresence system should enable an intuitive and simple interface to be made between the operator and the system. The visual display system should offer a field of view and resolution which is comparable to the operator actually being at the remote site in person. Ideally the mechanical system should have a zero latency period such that as the operator's head moves the mechanical system simultaneously follows the trajectory.
CHAPTER 3

STEREO PERCEPTION AND DISPLAY DEVICES

3.1 Introduction.

Due to its very large bandwidth the human vision system is one of the most important senses. It is the availability of this bandwidth which makes it very attractive for use as part of a teleoperator system. The vision system could be provided with a view of a remote scene through a single camera. However there is an inherent limitation with the direct use of single camera systems, namely the inability to directly determine spatial depth without the use of depth cues. By using cues such as object occlusion or perspective humans are capable of inferring the depth contained in the scene from a single camera. Unfortunately, there are limitations to this due to these depth cues being highly environment dependent.

An alternate and much more obvious method of realising depth is through binocular vision. Binocular vision relies on the principle that there are two image capture devices, such as the human eyes, which are separated by some distance and for the human eyes this separation is in the order of 65 mm (Hofsette 1972). In the case of the human vision system, a complete retinal image of the objects in the environment is formed in each eye. These two images are then fused by the brain to generate a three dimensional representation of the environment space and subsequently stereo visual perception is achieved.
Stereoscopic visual perception offers many advantages to the field of teleoperation and telepresence. By providing the operator with a three dimensional representation of the remote environment the feeling of presence is increased within the remote location and the perception of depth is enhanced. It has also been shown that when used with tele-manipulators that the time to completion of tasks is significantly reduced (Touris, Eichenlaub and Merrit 1992) and the accuracy is improved (Mclean, Prescott and Prodhorodeski 1994). The stereoscopic representation of the environment is often generated through the use of a two camera system, which simulates the presence of the human eyes in the remote environment.

This chapter aims to briefly review the physiological operation of the human eye, both in monocular and stereoscopic operation. It will also demonstrate the mathematics of stereoscopic camera systems utilising two cameras and review the various methods which can be used to display stereoscopic images.

### 3.2 Physiology of the Human Eye

Although the operation of the human eye is extremely complex, a basic understanding of the functionality can be gained by focusing on the physical properties of the eye. Very simply, the human eye consists of a cornea, an iris, a crystalline lens and the retina. The cornea is the transparent bulge at the front of the eye and acts as the eye’s first optical surface. Behind the cornea lies the anterior chamber which leads onto the iris. The iris is an extremely delicate membrane with a circular opening, known as the pupil, at its centre. The pupil dilates in accordance with the amount of light which enters the eye and in a normal human the pupil diameter varies between 2 to 8mm, a simplified schematic of the human eye is shown in figure 3.1.
Behind the iris is a crystalline lens which provides variable focusing in order to bring images to focus on the back of the eyeball known as the retina. The retina serves to convert the incident radiation into nerve pulses through its distribution of photosensitive receptors.

![Figure 3.1 A Simplified Schematic of the Human Eye.](image)

There are two groups of neural receptors on the retina, a rod system and a cone system. The rod receptors are unevenly distributed throughout the retina and are absent from a central area, which subtends approximately 1° of visual angle, known as the fovea. Rod receptors tend to react to light of a low intensity. Conversely, the cone receptors are highly distributed in the area of the fovea and are responsible for bright light, colour and visual sharpness or acuity. Both cones and rods are connected to nerve fibres which are connected to the brain where image perception occurs.

The area described as the fovea is the section of the retina which produces a high quality image and it has been estimated that for the entire retina to produce a high quality image would require the optic nerve to be increased in cross sectional area by a factor of two hundred (Carpenter 1988). It is for
this reason that, in mammals, high quality vision is limited to this central zone of the retina. However humans are very unaware of this limitation due to the eyes continually scanning the scene and this is known as scanning the field of fixation.

The rest of the surrounding retina provides what is known as peripheral vision. Although this is of lower resolution it can still localise on a particular area of interest and stimulate the brain to direct the fovea such that it may examined in more detail.

3.2.1 The Monocular Visual Field.

The ability of a human to see an object depends, firstly, whether or not it appears on the retina and, secondly, where on the retina it appears. The position of the object on the retina, as explained in section 3.2, determines the degree of resolution with which the object is viewed. Any object which is in close proximity to the eyes may not appear on the retina due to the field of view of the eye. The field of view of the motionless eye is very extensive, comprising of a horizontal width of approximately 155° and a vertical range of approximately 50° upwards and 70° downwards (Southall 1961). The horizontal range is not evenly distributed due to the protrusion of the nose and includes an angle of around 60° on the nasal side of the field of view.

Through scanning of the field of fixation, without head motion, it is possible to increase the peripheral field of view to yield a horizontal range of nearly 180°.
3.2.2 The Binocular Visual Field.

Humans have two eyes which are symmetrically located with an average interocular separation of 65mm (Hofsetter 1972). The two monocular fields of view overlap to form a region of overlap known as binocular overlap region where stereoscopic vision is achieved. The field of view of the binocular region is generally defined as the total field of view as seen by both eyes directed to a distant fixation point. The binocular field of view is used for central vision. With the eyes fixed in a forward looking position the binocular field of view is about 15° either side of the centre line and with eye motion this is increased to approximately 20° either side of the centre line. The field of view in the vertical plane is around 60° above the horizontal and 70° below (Kalawsky 1993).

Due to the separation of the eyes, two separate and slightly dissimilar monocular views of objects lying within the binocular overlap area are formed on each retina. It is these two views which are fused to form a single stereoscopic mental perception of the scene.

3.3 Stereoscopic Properties of Human Vision.

One of the properties of normal vision with two eyes is that, although there are two separate retinal images, for the most part only a single perception of the scene is experienced. The process of transforming two separate views into one single precept is known as fusion (Diner and Fender 1993). The fusion process involves several stages, the first of which is known as vergence. Vergence is the process by which both the eyes rotate to look at the object of interest. This causes the visual axes to converge towards the object and the point onto which the axes are converged is known as the point
of fixation. Simultaneously the eyes are also accommodated to focus onto the point of fixation where accommodation is the focusing action of the eye.

The next stage of the fusion process is to determine the correlation between the two retinal images. Due to the spatial separation of the eyes, the two images will not contain exactly the same information and so will not fit together exactly. This is due to some regions of the object or scene being represented in the left eye image but not in the right eye image or vice versa and consequently parts of the images are monocular. The observer actually sees a fused version of the scene, although a substantial portion of the image is not seen by both eyes.

The relative depth contained in an image is formed through a principle known as retinal disparity. If the eyes are converged on a single point on an object, say point A, in a scene then these points fall on exactly correspondent points in both eyes. Consequently the images of another point, say point B on a different plane, must fall onto non corresponding retinal points. It is the difference in location of points B and A on the retina which is known as retinal disparity.

This leads to the final stage of stereoscopic fusion where the disparate images are formed into a single model with the added property of relative depth. The disparity contained in the images is measured by disparity detectors in the eye, with an average resolution of 20 arc sec (Poggio and Poggio 1984).
3.3.1 Stereoscopic Acuity and Stereoscopic Range.

The stereoscopic acuity of an observer is a measure of when the observer can just distinguish that two objects in space are lying at different depths. This is limited by the resolution of the disparity detectors in the eye. This value of stereoscopic acuity also sets a limit on the furthest distance beyond which the individual cannot detect differences in depth.

If the eyes are converged on an object at a distance D, then the angle \( \theta \) between the two visual axes is given by the equation:

\[
\frac{\text{IPD}}{2D} = \tan \left( \frac{\theta}{2} \right)
\]

where IPD is the inter ocular distance between the two eyes. Therefore, as the disparity angle tends towards the limiting value of 20 arc sec, the limiting value at which stereoscopic depth can be sensed can be determined. So for an average inter ocular separation of 65mm, the limiting range can be found to be 670m. Thus for a given observer any object at a distance greater than 670m will not be separable from one at infinity, using only stereoscopic vision as a depth cue.

3.3.2 Human Visual Stereoscopic Space.

The relative depth of objects which are viewed in an environment is determined by the disparity which is experienced on the retina of the eyes. When the eyes exhibit a vergence motion to focus on a particular object, the object is perceived as a single object and presents zero disparity. This is also true for a single feature on a given object. It would therefore seem practical
to assume that any other points on the same depth plane would also exhibit zero disparity, however this is not the case.

If the eyes converge on a given object, $O_1$, in the environment, figure 3.2, then the images exhibit the same disparity in both retina and are fused to form a single object. If it is assumed that the eyes are identical in shape and size then it can be shown from elementary geometry, that the locus of points which exhibit the same retinal disparity, $\delta$, is of a circular form. All points which are in front of the eyes and at which $\theta_1=\theta_2$ must lie on a circle which passes through the points $O_1$ and $O_2$, this is known as the Vieth-Muller Circle.

![Diagram of Vieth-Muller Circle](image)

Figure 3.2 The Vieth-Muller Circle.

As the images $O_1$ and $O_2$ are fused by the observer and since the images of $O_2$ display the same retinal disparity they will also seem to be fused. Consequently point $O_2$ will appear to be located at the same depth in space.
It is also true that an object, $V$, located on the plane of convergence will appear to the observer to be further away than $O_2$ and subsequently further away than $O_1$. The effect of this is that binocular space is in fact curved convexly away from the observer and is highly apparent in environments which are devoid of many other types of depth cue, such as space.

Through using elementary geometry regarding cords of circles and a consideration of figure 3.2 an assessment of the magnitude of the curvature of stereoscopic space can be evaluated. Equation 3.2 allows the curvature at a point of convergence, i.e. $O_1$, on the centre line to be calculated.

$$c = \frac{1}{r} = \frac{2v}{(v^2 + \frac{I^2}{4})}$$

(3.2)

where $v$ is the distance to the plane of convergence and $I$ is the interocular separation. So for $V$ equal to 1m and $I$ equal to 65 mm, the curvature at the point $O_1$ can be shown to be 2m$^{-1}$.

The sensation of curved binocular space is very often overwritten by contextual or stored mental evidence regarding how a scene should appear. Consequently the curvature of binocular space can be sensed to be minimal and also have minimal effects in everyday life. However, it is of interest in the generation and understanding of the environment which is viewed by an observer through a 3D viewing system.

3.4 Stereoscopic Camera Systems.

Although not very popular or commercially available, stereoscopic camera systems can theoretically be formed by utilising a single camera. A single camera system requires the use of numerous mirrors which must be
accurately aligned to prevent rotation or skewing of the images taking place. Skewed images cause a false disparity to be formed and the images appear to lean in a particular direction (Diner and Fender 1993). For this and other reasons, such as non-linearities present in the mirrors, double camera systems are the preferable format for stereoscopic camera systems.

Double camera systems enable two viewpoints to be obtained from slightly different positions. The typical configuration of these systems enables a number of analogies to be made with the human vision system, such as the interocular distance and the vergence angles. Much of the supporting mathematical theory of two-camera systems is very much independent of the type of cameras which are being used.

There are a number of properties of two-camera systems which enable comparisons to be made between different systems and spatial configurations, examples of which are (i) stereoscopic magnification and (ii) magnified stereoscopic depth. The stereoscopic magnification of a system can be determined by considering the schematic shown in figure 3.3.

The two cameras are focused on an object, \( O_1 \), which lies on the perpendicular bisector of the two lens nodal points, \( N_l \) and \( N_r \), such that the camera images are formed at \( O_{1l} \) and \( O_{1r} \). If a display system is used which allows the observer to fuse the images \( O_{1l}, O_{1r} \) and an additional object, \( O_2 \) is placed a distance \( \Delta \) in front of \( O_1 \), then this will form images \( O_{2l} \) and \( O_{2r} \) on the camera.
These images will then be viewed with a disparity of $2\delta$. The stereoscopic magnification is given as the ratio of the disparity to the distance between the two objects or by

$$M_s = \lim_{\Delta \to 0} \frac{2\delta}{\Lambda}$$

By using elementary optics (Carlson 1977) it is possible to show that

$$\delta = \frac{f_c L}{L - f_c} \tan \phi$$

where $f_c$ is the focal length of the camera lenses and $L$ is the distance between the convergence point and the front nodal points of the lenses, $N_1$ and $N_r$. This then leads to a more useful form of the equation for stereoscopic magnification, given by equation 3.3.
From inspection of equation 3.3 it is apparent that for a given camera to object distance and fixed camera parameters, the stereoscopic magnification increases as the inter viewpoint separation of the camera pair is increased. As the stereoscopic magnification increases the magnitude of the perceived stereoscopic depth is increased. The depth effect becomes easier to detect but is more difficult to accurately estimate the magnitude of the depth between two points.

3.4.1 Stereoscopic Space - As Perceived by Camera.

The apparent curvature of stereoscopic space has been highlighted for the case of direct human vision. A similar situation occurs with images which are viewed by cameras, although some of the details are slightly different. The main difference between the two systems is in the form of the device which is used to collect the image data. The imaging surface of cameras tend to be planar as opposed to the assumed spherical shape of the human retina. A similar analysis can be performed to that on human vision, but, any use of small angle approximations will introduce errors into the calculations, due to flat camera plates. After rigorous geometrical analysis it can be shown that the stereoscopic space which is generated by the use of cameras does in fact consist of a family ellipses and not circles as in the case of direct human vision (Diner and Fender 1993). This does have implications for a remote presence system in environments with very limited depth cues.
3.5 Stereoscopic Display Devices.

There are many techniques of presenting stereoscopic images to an observer. These images may be static in form, such as photographs or prints, or they may be real-time images which have been generated from two television cameras or computer software. A good example of static 3D presentation are the now popular, and widely available, stereo-grams where a series of dots are viewed and when fusion occurs an image appears in great depth on the paper. This type of stereoscopic presentation does not lend itself to many practical applications as the images are very often difficult to visualise and are of course static, hence are more suited to entertainment.

More specifically of interest are the techniques of presenting real-time stereoscopic images to an observer. This can be achieved in many different ways using numerous different types of display technology. However, all the systems operate on one of two principles either the left and right eye images are separate and distinct, or the images are combined. Each of these display techniques are used in immersive or non-immersive displays. Non-immersive displays usually involve the operator viewing some form of television monitor whilst also being able to view their surrounding environment. Whereas, when using an immersive display system the operator cannot view the surrounding environment due to their visual field being totally encased by the display and supporting material. This type of display is generally described as a Head Mounted Display or HMD.

A number of different display techniques and systems will be discussed and the technical features of the more important systems will be elaborated upon.
3.5.1 Non-Immersive Separate Image Systems.

Separate image systems function by presenting separate left and right images to the left and right eyes. One example of this technique is where the images are displayed on two monitors which are located side by side. The monitors are then viewed through optics incorporating prisms. The observer's eyes then converge on the plane of the image pair to form a fused stereoscopic image.

An alternative method of presenting two separate images is to use a 50% transmitting mirror (Kusaka 1992). Two monitors, oriented at 90° to each other, are used to present the separate images and the half mirror is used to superimpose the two views. Each of the monitors has a polarising filter mounted over it, and the observer wears polarising glasses. The polarising filters on the monitors and in the users glasses enables only the right image to reach the right eye and the left image to reach the left eye, thus creating the stereoscopic effect. The major disadvantage of using this type of system is the difficulty of adjusting and maintaining the exact alignment of the two monitors and the half mirror.

Due to the inherent difficulties in the setting up of separate image systems, their use has remained very low in telepresence applications. Their overall size tends to be a lot larger than a similar system utilising combined imaging.

3.5.2 Non-Immersive Displays with Combined Images.

There tends to be two standard techniques for combining a pair of images to form a single stereoscopic view. These techniques are known as spatial overlay and temporal sequencing.
Temporal sequencing presents the stereoscopic pair, or left and right images, sequentially to the left and right eye. If this sequencing can be achieved at a high enough rate then the human vision system will amalgamate the two images into a single perception of space.

Spatial overlaying techniques can be used with both stationary and real time images. The principle operates by simply overlaying the left and right images on the screen. The screen is then viewed through an optical viewing device which will separate the images.

3.5.2.1 Spatial Overlaying Techniques.

Spatial overlaying is a very popular technique in stereoscopic cinematography (Lipton 1982) and usually assumes one of two forms. The first form of this display technique is similar to the principle of anaglyphs.

Anaglyphs are formed by projecting the left and right images in different primary colours, generally red and green. If the left and right images are projected through a red and green filter and the observer wears a similarly arranged pair of filter glasses. The green image will be visible when viewed through the green filter and the red image will appear black through the same filter, and the converse is true for the red image. Thus each eye will view the appropriate image, which will be fused together to generate a stereoscopic effect.

More recently, in cinematography, there has been a tendency to use polarising filters to produce the image separation. In a similar way to anaglyphs, the observer views the projected imagery through a
corresponding set of polarising filters and fuses the images to generate the sensation of depth.

An additional method which is sometimes used in cinema, although not essentially a spatial overlaying technique, is known as the Pulfrich effect (Pulfrich 1922). The Pulfrich effect can only be experienced with moving objects within the scene being viewed. It requires a polarising filter to be worn over one eye, the right eye for example, and the motion to be captured moving from left to right. This was used to good effect during the intermission of the 1988 Superbowl, which featured marching bands filmed such that they appeared to be moving from left to right. This then generated a sense of depth in the scene.

3.5.2.2 Temporal Sequencing Techniques.

The pictures for television systems are not transmitted complete, that is a single screen or frame is not formed in a single step as this would require significant bandwidth. Additionally it would be very wasteful to transmit 50 frames per second (for the PAL system) to avoid a flickering effect, when 25 frames per second are adequate to prevent flicker (Trundle 1992). Instead each frame is transmitted in the form of fields, which is known as field sequential television, and each frame requires two fields, one for the odd lines and one for the even lines. It is the availability of the different fields which is made use of during temporal sequencing.

During temporal sequencing the image for the left eye is displayed on the odd lines and the image for the right eye is displayed on the even line field of each

1 PAL - Phase Alternation Line.
frame. In order for each eye to view only the relevant video image it is necessary to shutter or polarise the images and this is achieved in one of two ways.

The first method uses electro-optical shutters, one for each eye of the observer. The shutters are switched to be either transparent or opaque out of phase with each other. The switching is synchronised to the field rate of the monitor, such that each eye only sees the appropriate image in time sequence.

The second method utilises a polarising liquid crystal display in front of the monitor screen. On application of a biasing voltage the plane of polarisation can be varied. The biasing voltage is synchronised with the field rate of the monitor and through the observer viewing through a pair of polarising glasses the two, left and right fields, can be viewed sequentially. Initially this type of system used linear polarisation but during any observer head motion the stereoscopic effect would be lost. Consequently the linear polarisers were replaced by circular polarisers which give an improved stereoscopic effect during any observer head motion.

When a PAL monitor is directly viewed in monoscopic form, each eye sees all 50 presentations per second. However, during temporal sequencing each eye only sees 25 fields per second and so it is possible for perceived flicker to occur. To ensure that the display is totally flicker free, it is possible to double the field rate.

The temporal sequencing method of stereoscopic viewing offers one of the clearest and flicker free approaches to non-immersive stereoscopic viewing.
The limitation is that the liquid crystal display used for polarising the left and right images tends to be very expensive.

3.5.3 Immersive Display Systems.

Immersive display systems usually take the form of Head Mounted Displays (HMD), that is the display is mounted directly in front of the observers eyes in a helmet type of attachment. HMDs consist of two display sources and a series of optical elements which serve to project the image at some distance in front of the observer. The images appear to exhibit a degree of overlap, which is dependent on the HMD design, which enables the images to be fused and thus generate a three dimensional image in the binocular overlap region. The optics also serve to magnify the image so that it fills a large part of the observers field of view.

Many different types of technology can be used to display the images, but the most popular are cathode ray tubes (CRTs) and liquid crystal displays (LCDs). In addition these two technologies have now been combined to form colour shutter CRTs, which offer significant advantages, in resolution terms, over pure CRT based systems.

The availability and relatively low cost of LCDs has meant that their use within HMDs has become very popular in recent years. They offer reasonable resolution providing that the field of view is not too large because of the limited resolution of the LCD. Higher resolution LCDs, of around 416 by 277 RGB pixels (VPL HRX Helmet), have been used with an associated increase in cost.
If LCD and CRT display technologies are combined a colour shutter CRT display is formed. The LCD device is placed after a monochrome CRT and is used to select the colour which is to be displayed. Early versions of this technique could only allow colours in the range of red to green to be reproduced. However, this did offer significant advantages as the spatial resolution of the monochrome CRT was maintained.

Subsequent developments in the technology have enabled LCD colour shutter displays to be capable of forming full colour high resolution displays. This overcomes many of the problems which are experienced with attempting to produce miniature colour CRT devices. The difficulties arise because of the need to generate the three primary colour images, red, green and blue, which are required to produce a full colour image, on a small area. It was evident that using current fabrication techniques that this could only be achieved if there was a significant decrease in the available resolution.

Early attempts to overcome these problems involved spinning a coloured filter wheel in front of three monochrome images, one for the red, green and blue components. This method produced good results, but unfortunately due to its mechanical design suffered with reliability problems. Tektronix, an American company, developed an alternate approach to the idea of a colour filter wheel. The wheel is replaced by a colour polarising liquid crystal shutter, which splits the red, green and blue components of the CRT emission into orthogonal polarised components. In simple terms the CRT is made to outline the red, green and blue components of the scene individually and the liquid crystal filters generate the respective colour content. In order for the display to be flicker free and compatible with standard video systems the shutter must be switched at three times the primary field rate, which is 50 Hz for the PAL system.
The use of LCD colour shutter systems in HMDs has enabled very high resolution systems to be developed achieving resolutions of 1024 x 1024 in full colour from a 1 inch CRT. However, the significant increase in performance is not achieved without an associated increase in price, at current prices a Tektronix LCD colour shutter CRT helmet is in excess of fifty thousand pounds. This is approximately ten times the cost of a lower resolution LCD based helmet.

Due to the lack of availability of miniature CRTs with high resolution and full colour, alternative techniques of using larger scale CRTs have been developed. These have generally relied on the use of fibre optic technology (Shenker 1987). A system developed by CAE of Canada utilised high resolution CRT displays from which the images were directed to the observer by way of flexible fibre optic bundles. These systems do provide an alternative high resolution solution to the very expensive LCD shuttered HMDs, but they do require two rather large bundles of optical fibres to be connected to the helmet.

3.6 Summary.

This chapter has outlined the processes by which the human visual system formulates stereoscopic views of the environment. It also demonstrates that human stereoscopic visual space is non-linear and exhibits a degree of curvature. However, due to the existence of alternative depth cues such as occlusion and perspective the curvature of visual space is not experienced during normal everyday vision. Further more, the stereoscopic space viewed by verged cameras exhibits a slightly different form to that of human vision, it is in fact parabolic. These effects can force objects which are at the corners of the images to appear to be further away than objects at the
centre of the image. This would have significant effect when used in environments which are devoid of many other forms of depth cue, such as in space related applications.

Many of the state of the art and commercially available techniques for displaying stereoscopic images have been reviewed. This provides an insight into the complex techniques which are used for stereoscopic viewing and highlights the differences and shortcomings of many of the systems.

Through the demonstration of the mathematics of stereoscopic vision systems possible methods for evaluating stereoscopic camera systems, such as stereoscopic magnification, have also been developed.
4.1 Introduction.

There are many essential attributes of a telepresence system; it must consist of a sensor system which gathers information from the remote environment and a suitable method of displaying this information to the operator. In addition to the presentation of the visual information, the operator must also be capable of interacting with the remote system. This must be achieved in such a way that the operator does not feel visually separate or removed from the remote environment. The images presented to the operator may convey a sense of being present at the remote environment, but when the operator moves their head the feeling of presence will be reduced. In order to overcome this it is proposed that the system should be slaved to the operators head position thus providing an enhanced feeling of presence. Additionally during the process of head slaving the system must be capable of responding to the dynamics of the operator’s changes in gaze direction.

The design of a high performance telepresence system requires the integration of many different aspects of mechatronics. The dynamics of the system must be capable of exceeding the performance of the human operator, so as to not introduce any operator disorientation, which must also be controlled in real time. A sensor system must be used to track the motion of the operator’s head which must then be integrated with the
actuation controller. The image collection devices and the display systems must form an integrated sub-system of the telepresence system.

All of the systems highlighted require careful consideration and place varying demands and constraints on the system design. The main objective of the design considerations is to analyse and evaluate the requirements of all the various sub-systems from which the telepresence system is comprised, which will enable a high performance telepresence system to be developed. The chapter aims to suggest suitable solutions for the head tracking sensor system, the display systems and an overall view of the control system architecture. In doing this, it doesn’t aim to describe how the individual components are interfaced together to form the integrated telepresence system, but it will identify suitable solutions.

4.2 Mechanical Design Requirements.

The mechanical design of the telepresence head is of paramount importance as it determines the functionality of many other aspects of the system. The mechanical design parameters can be classified into two different categories (i) the static and (ii) the kinematic requirements.

4.2.1 Static Mechanical Design Considerations.

The static mechanical design is involved with the evaluation of the mechanical properties of the head such as the interocular separation, the overall weight of the system and joint stability. The remote head must occupy a minimal space envelope not exceeding a cylinder of dimensions 300 X 300mm diameter and exhibit a total mass of less than 10 Kg. The limiting weight and space envelope was chosen as this is the maximum
payload and the overall dimensions of a small laboratory mobile robot onto which the system was to be initially mounted.

As explained in Chapter 3, the formulation of stereoscopic imagery requires that the images are formed from two disparate points. This separation of the cameras is known as the inter ocular separation of the remote head. In order to account for future experimentation, it was desirable for the inter ocular distance to be manually configurable over a range of 50 to 150mm. This range enables experimentation to be performed using separations both greater and smaller than the average standard human eye separation of 65mm (Hofsetter 1972).

During power down of the system it is desirable that the system will not simply collapse causing damage to itself or the surrounding environment. The axes which, when power is removed, would be unstable are required to be either mechanically braked or exhibit a high degree of natural damping to prevent any damage occurring. It would be preferential, in design terms, for the system to be naturally balanced such that no external braking system is required as these are generally costly and rather bulky.

4.2.2 Kinematic Design Considerations.

The first stage in assessing the kinematic design for a telepresence system is to consider the formation of the human ocular and head motor system motions. The motion of the human neck and ocular system are very complex. The human neck consists of many muscle and bone interactions which are capable of generating complex head motion trajectories including rotational and translation motions. The translation motion of the neck is very minimal and will not be considered in the kinematic model to be
developed here. Similarly the human eye is capable of the generation of complex trajectories. Both the eyes and the head have numerous degrees of freedom, but for analysis of the motions these will be reduced to consider only the rotational motions of each joint. The neck and each eye are considered as single joints each with three degrees of freedom and are represented in the schematic of figure 4.1.

The additional degrees of freedom which are exhibited by the eyes are not shown in figure 4.1. These additional motions, such as focusing and iris control, are accounted for in the camera system and so can be viewed as not a definitive requirement of the mechanical mechanism.

![Diagram of human neck and eyes rotational degrees of freedom]

Figure 4.1 Simple Model of the Human Neck and Eyes Rotational Degrees of Freedom.
The dynamics of the head and eye system exhibits a certain degree of redundancy in the elevation direction. It is essentially possible to achieve the same direction of gaze or viewpoint by an elevation motion of the eyes or of the neck and indeed during normal visual activity these two axes function in collaboration. The main difference in the motion capable by rotation about these two axes is the range of motion which can be achieved. During neck elevations the eyes move on an arc due to a certain degree of offset between the centres of rotation, hence a greater volume of the workspace can be viewed for a given angular movement. Although not identically the same kinematically, the effect of the elevation motions of the eyes can be simulated by the elevation motion of the neck. Consequently the elevation motion of the eyes is not modelled in the mechanical system.

The maximum roll movement of the eye which is exhibited by most people is of a very low magnitude, and has been demonstrated to be around 7° (Collewijn, Ferman and Jansen 1985). Research has shown that the torsion of the eye is a vestibular reflex action which normally only occurs during roll of the head as an attempt to compensate for the angular orientation of the head. Due to the relative small angular magnitude of motion required by the roll movement, it was deemed unnecessary to include it in the mechanical system. It may, however, prove necessary to be included in the design of future advanced systems.

Both of the eyeballs exhibit a panning motion. This motion is utilised during vergence of the eyes and should be incorporated into the mechanical system.

The motion of the human neck has been reduced to three orthogonal components, pan, elevation and roll. Both the pan and elevation motions
enable a visual search of the workspace to be performed and allows the apparent field of view to be increased. Hence these two axes of motion form an essential aspect of the mechanical arrangement of the telepresence system.

The third degree of freedom of the neck is roll. During most teleoperated tasks the necessity to perform a roll motion of the neck would be expected to occur relatively infrequently. This is due to the motion not presenting any new information to the operator. It merely acts to enable the sensation of realism to be maintained between the operator and the mechanical system. However, the inclusion of this degree of freedom in the mechanical structure would make the mechanism more complex and also require a further actuator, thus increasing the cost. The decision was therefore made not to include the roll degree of freedom in this generation of the telepresence system.

These considerations lead to the formulation of a reduced degree of freedom model of the human neck and eye system, figure 4.1b. The system is now modelled using four degrees of freedom; pan and elevation of the neck and the pan motions of each eye, instead of the original nine.

4.2.2.1 Range, Velocity and Acceleration of Each Degree of Freedom.

A critical aspect of the performance of a telepresence system is the ability of the mechanism to simulate the human head motions. This must not only be achieved in directional terms but also in terms of the range, velocity and acceleration which are demanded by the operator. The mechanical system must therefore be capable of at least matching the dynamic performance of
the human operator. If this is not the case the system will appear to move much slower than the operator, thus generating a disorientating sensation and lack of operator telepresence. In addition, if the operator can move to positions which the mechanical system cannot reach, this will also introduce disorientation. Through limiting the range of motions which can be achieved, the operator is highly constrained and must perform extra work to view areas which would naturally be visible.

The ranges of motion which are generally achieved by a human neck are \( \pm 60^\circ \) and \( \pm 45^\circ \) for the pan and elevation axes respectively (Gowitzke 1980). The two rotational motions of the eyeballs enable the eyes to verge towards an object and generate stereoscopic fusion and for symmetrical version it is possible to rotate the eyes through an inward angle exceeding \( 45^\circ \). A range of at least \( \pm 45^\circ \) is thought to be sufficient for the vergence mechanisms.

As previously stated, each degree of freedom must be capable of matching the rotational velocity and angular acceleration of the human in order for the sensation of telepresence not to be compromised. There have been many studies performed in order to evaluate the rotational velocity and acceleration of the human head with a range of results a summary of which are shown in table 4.1 (Bolas and Fisher 1990, Carpenter 1988, Du and Brady 1992).

Table 4.1 shows the dynamic performance requirements on which the range, velocity and accelerations of the mechanical system are to be based.
Chapter 4: Mechatronic Design Considerations.

Table 4.1 Summary of Mechanical Head Kinematic Requirements.

<table>
<thead>
<tr>
<th></th>
<th>Range °</th>
<th>Velocity °/sec</th>
<th>Acceleration °/sec²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pan.</td>
<td>± 60</td>
<td>400</td>
<td>2000</td>
</tr>
<tr>
<td>Elevation.</td>
<td>± 45</td>
<td>150</td>
<td>1000</td>
</tr>
<tr>
<td>Vergence.</td>
<td>± 45</td>
<td>400</td>
<td>4000</td>
</tr>
</tbody>
</table>

4.3 Vision System Requirements.

The vision system can be separated into two subsystems (i) the camera system and (ii) the display system. The cameras are mounted on the mechanical mechanism, whereas the display devices are located near or on the operator.

4.3.1 The Camera System.

The camera comprises of two components, the pick up device and the lens. Nowadays the majority of cameras which are suitable for telepresence or machine vision applications are of the Charged Coupled Device (CCD) type. These are available in a vast majority of different formats with correspondingly different operating characteristics.

To maintain the light weight and highly responsive characteristics of the telepresence system the cameras need be light weight and compact. Miniature television cameras fall into two distinct types. One type maintains a small space envelope by having a remote electronic head unit with the actual physical camera section only containing the imaging chip.
and some associated electronics. This enables the camera unit to be extremely lightweight and compact. The other type of camera incorporates the controlling electronics local to the imaging chip, thus forming a larger and heavier camera unit. It was therefore preferable to only include miniature camera systems which utilised a remote head unit, although they are significantly more expensive.

The cameras must enable standard PAL or S-VHS colour video signals to be produced as these offer an easy interface to additional video equipment. In addition to being colour, the camera's must have genlocking capability, low image smear and high resolution as described below.

Resolution of a video camera is a measure of the ability of the imaging sensor to distinguish detail in the image. A consideration of the camera resolution is very important in telepresence applications. If the camera is of very low resolution the images would appear to the operator to be of poor quality and hence be detrimental to the sensation of presence. An ideal solution would be to match the resolution of the camera with that of the human eye. However with current technology this is not a realistic possibility. The human eye receptors have a separation of around 7µm (Jenkin and Tsotsos 1992) in the area of the fovea whereas the current generation of commercially available high resolution cameras have a pixel separation of around 10µm. The resolution of the camera must therefore be as high as the current commercially available colour CCD cameras can achieve.

The requirement of the cameras to exhibit the capability to be genlocked is a paramount consideration when utilising multiple camera systems. In its simplest form genlocking allows the field and line scans of multiple
cameras to be synchronised. During genlocking the synchronisation of one
camera is locked with that of either another camera or the master
synchronisation source or ‘sync-generator’. The ‘sync-generator’ supplies
the vertical, horizontal blanking and other signals which are required for
the camera operation. There are several methods of synchronising multiple
video signals such as Phase Lock Loop (PLL), direct clock drive and
through the use of an internal clock (Martins 1988). Direct clock drive is
the simplest method to implement as the synchronisation is achieved by an
external clock input such as the main clock of a computer along with the
horizontal and vertical driving information. Genlocking of multiple image
sources is very important for the generation of stereoscopic video using field
sequential display techniques (Section 3.5.2.2).

It must be possible for the camera system to be capable of generating good
quality images in a variety of environments. It may be probable that during
a particular application the camera system will move from an indoor
artificially illuminated environment to an outdoor naturally illuminated
environment. A change in environment such as this can entail an increase
in light levels by a factor of 10000 lux (Trundle 1992) and humans
accommodate for this change in light level by reducing the aperture size of
the iris which is analogous to the operation of the auto-iris of a camera lens.
It is now possible to replace the mechanical auto-iris by electronic feedback
control known as auto gain compensation (AGC). AGC serves to modify the
gain of the imaging system to prevent saturation occurring, thus the
sensitivity of the CCD chip is modified according to the ambient lighting
conditions. This offers a significant weight and space saving over the use of
mechanical auto-iris mechanisms which tend to be rather bulky while
offering a similar performance.
Aspects of image smearing must be considered due to the induced motion of
the camera system when mounted on the camera platform. The relative
motion between the camera and the scene can cause the images to become
blurred or smeared, dependent on the rate of motion. The generation of
smearing of the image can be reduced by using a camera with a higher
shutter speed, that is one which can accumulate more images per second.
Some modern cameras enable the electronic shutter speed to be manually
configurable in a range of 1/50sec to 1/10000sec, thus enabling the
reduction of image smear at high rates of motion if necessary.

The remote head type of camera, in general, do not have a
universal/standard lens connection. The lens connection tends to be very
manufacturer specific, but if necessary a standard C-mount adapter is
usually available.

### 4.3.1.1 Camera Lens Considerations.

There are a number of considerations which must be made in order to
evaluate the most suitable type of lens for the telepresence system. These
include, the field of view, depth of field, focusing mechanism, near point,
space envelope and weight.

For an ideal telepresence system the field of view which can be attained by
the camera lens should match the human field of view as closely as
possible. The human field of view extends to approximately (Section 3.2.1)
155° in the horizontal plane, which in terms of camera lenses is deemed to
be wide angle. The difficulty in realising the human field of view with
commercially available lenses is that very wide angle lenses tend to
introduce significant distortions into the images and a corresponding
reduction in resolution, which is demonstrated by the apparent curvature of a scene taken by a camera utilising a ‘fish-eye’ lens. Distortion of this kind is not acceptable in a telepresence system as it yields a very unnatural view of the remote environment. The problem of distortion in wide angle camera lenses has been approached by the development of Large Expanse Extra Perspective lenses by LEEP Systems, USA (Howlett 1992b). These lenses are noted to achieve a horizontal field of view of 140° with little distortion, providing similar lenses are used in the viewing device (Bolas and Fisher 1990). However these lenses are extremely expensive, around £4000 for the camera and viewing lens combination. It has therefore been decided to limit the selection of lenses to those available ‘off the shelf’ and suitable for use on miniature remote head type of camera without the use of adapters.

The depth of field, near point and focusing mechanism are very much related to each other in the consideration of the type of lens to use. The focusing mechanisms of lenses fall into two categories, manual and motorised. Motorised focusing of lenses tends to make the lens heavier and larger and are in some cases difficult to control (Sharkey et al. 1992). The depth of field of a lens determines the depth of the image scene which is in focus, that is a larger depth of field the more of the scene is in focus. So, if a lens with a large depth of field can be configured such that a large degree of the scene is in focus then the requirement of motorised focusing is somewhat reduced. However, there are other factors which must be considered for the suitability of this technique. If the depth of field is such that any object within 1m of the lens is not in focus then this would obviously be unacceptable for remote manipulation applications. It is possible to utilise a commercially available lens which has a near point of approximately 0.2m and a reasonably large depth of field. Therefore through a compromise of the various requirements a suitable lens configuration can be determined.
such that the near point, depth of field and field of view combination can be deemed to be acceptable.

The field of view should, therefore, be as wide as possible without the introduction of geometric distortion or excessive cost (Bolas 1994). The lens should demonstrate a near point of 0.2m, which is an approximate human manipulation distance, and be directly compatible with remote head cameras.

4.3.2 The Display Systems.

As detailed in section 3.5, there are essentially two types of stereoscopic display device, immersive and non-immersive systems. The telepresence system is proposed to utilise both methods of display system to allow a comparison to be made. The immersive system is to be used by the operator to whom the camera system is slaved and the non-immersive system can also be used by other external observers for demonstration purposes.

There are many different types of Head Mounted Display (HMD) systems currently available on the market and range in price from £1000 to in excess of £50000 (at current market prices). As expected the quality and resolution of the display is highly price dependent and at the low end of the stereoscopic display market the displays exhibit similar characteristics of resolution and display quality. Due to the American driving force behind Virtual Reality (VR) in the past, the majority of these HMDs require NTSC\textsuperscript{1} video inputs and so it is necessary to utilise a PAL to NTSC converter. A reasonably priced and easily available HMD is the Eyegen3 helmet. The Eyegen3 helmet was developed by Virtual Research and tends

\footnote{National Television Systems Committee.}
to be one of the better HMD in the £5-10k price range. The helmet uses dual monochrome CRTs with colour wheels instead of LCDs and requires two composite NTSC video signals. A summary of the Eyegen3 specification is given in table 4.2.

The choice of good quality non-immersive commercially available display systems is very limited of which the leading systems are manufactured by Stereographics and Tektronix. These two systems operate on the same principle of temporal sequencing (Section 3.5.2.2). The only differences being that the Stereographics system relies on the viewers wearing LCD shuttering glasses, and the Tektronix system relies on the viewers only wearing glasses with circular polarising lenses. The system performances are very comparable and the underlying advantage of the Tektronix system is that the number of people who may view the stereoscopic display can easily be increased by purchasing more of the very low priced circular polarising glasses as opposed to the very expensive LCD shutter glasses. The choice of non-immersive stereoscopic display was therefore confined to the Tektronix display system.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>493 X 250 resolvable lines.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Signals</td>
<td>Composite NTSC.</td>
</tr>
<tr>
<td>Optics</td>
<td>Horizontal</td>
</tr>
<tr>
<td>Field of View</td>
<td>40°</td>
</tr>
</tbody>
</table>

Table 4.2 Eyegen3 Specification Summary.
4.4. The Head Tracking Sensor System.

A sensor system is required to provide real-time position sensing of the operator's head. The data generated by the sensor is used as the demand signal for position control of the system. The sensing device should allow both position and, more importantly for this application, should enable the orientation of the operator's head to be reported.

The performance of the head tracking system is very important in terms of achieving convincing telepresence, as the gaze direction of the camera system is directly linked to the perceived sensor position. The performance of position tracking devices can be evaluated by considering the accuracy, resolution, update rate and system latency. A good position tracking system should have a high accuracy, fine resolution and a high position update rate. Approximate acceptable values for these parameters are; resolution < 0.5°, accuracy < 0.5° and a position update rate in excess of 100Hz (Kalawsky 1993). The latency of a system is the delay between the change in position of the sensing device and the reporting of this change to the external system. Although it is not technically possible to eliminate latency, it should be as low as possible. The effect of system latency on a telepresence system is very similar to the effect which is experienced with a Virtual Reality (VR) system. It can lead to a sensation of disorientation, for example the operator may be beginning to look to the left but the telepresence or graphics system has yet to respond. Obviously, the sensor latency is only part of the overall system latency, which in a telepresence system also comprises of effects due to the controller performance (Bryson and Fisher 1990). So, ideally the sensor system should exhibit a minimum lag in order to reduce the total latency time. Experiments have been performed (So and Griffin 1992) which have shown there to be a significant
performance degradation with lags greater than 100ms and a latency of this magnitude or greater can induce disorientation or motion sickness effects. This particular experimentation was performed using VR equipment but is also relevant to telepresence systems.

Head position tracking systems can be broadly classified into four main categories, (i) magnetic, (ii) acoustic, (iii) optical and (iv) mechanical. The first three types of sensing systems are non-contact systems, that is, there is not a direct physical linkage between the head and the sensor. Each of the sensing systems have their own advantages and disadvantages.

Acoustical systems tend to use ultrasonics to determine the position of the target object, which in this case would be the operator's head position. In general these are based on calculating the time of flight of ultrasonic pulses. A time of flight system could consist of three ultrasonic transmitters located on the object to be tracked and three receivers located at known positions within the environment. At regular intervals, and in turn, each transmitter emits a pulse which is detected by the receivers. The time of flight of the pulses can be measured and hence the distances between the transmitters and receivers calculated. Through triangulation the position and orientation of the transmitters can be evaluated (Ferrin 1991).

The main advantage of the ultrasonic tracking system is that they are unaffected by the presence of magnetic fields in the environment. However they are unable to function correctly if there are obstructions between the transmitters and receivers. They can also suffer from inaccuracies due to any echoes or noise in the surrounding environment and also have a low position update rate, typically in the region of 50Hz. Time of flight position tracking systems are not very suitable for use in high performance
telepresence systems as their update rates are low and the latency times are very long, typically in excess of 30ms, in comparison with other forms of sensing. They also require considerable configuration of the receiver systems.

Optical trackers utilise a wide variety of differing technologies and can generally be classified as either beacon based or laser ranging based trackers. Beacon based tracking systems use a set of optical beacons and a set of sensors to track the beacons which may be either passive or active. An example of an optical system is that developed by Cook of NASA (Cook 1988), which uses four cameras to track the position of a set of LEDs which are mounted on the operators helmet. The system uses the measured position of the images in the cameras to compute the head position and orientation.

Laser ranging tracking systems utilise many different techniques to evaluate the position and orientation of the object being tracked. One method utilises diffraction gratings and image processing techniques to determine the object orientation and position.

The performance of optical based tracking systems is very much limited by the speed of the processors which are used to perform the, often, complex image processing algorithms. However position update rates of around 80Hz (Ferrin 1991) are achieved in some systems. Optical tracking systems suffer from the same line of sight problems which are experienced with acoustic tracking systems. Again, due to the fact that the position calculations involve triangulation, the beacons must be accurately arranged within the environment. An often useful feature of optical trackers is that they do not suffer from distortion effects due to the presence of metallic
objects, as do magnetic systems, they are however susceptible to effects due to the ambient lighting conditions.

Mechanical tracking systems measure the operator’s head position and orientation through the motion of a mechanical arm. The arm is anchored at a fixed point and consists of a number of sections and joints which can rotate and move. The position and orientation is calculated through a knowledge of the arm forward kinematics and the joint positions. The joint positions are usually measured using potentiometers which are interfaced to the controlling processor by means of analogue to digital converters (ADC). An example of such a system is known as the ‘BOOM’ and was developed by NASA (McDowall et al. 1990).

The mechanical linkages are generally counter balanced such that the operator is not subjected to excessive loading and the HMD usually forms part of the integrated system. The advantage of a mechanical tracking system is that the processing to evaluate the position and orientation is very simple and can so be completed very quickly, thus increasing the update rate and reducing the system latency. However a major disadvantage of mechanical tracking mechanisms is that the user is constrained by the mechanical arm and does not have full freedom of movement. The working volume is also quite restricted due to the reasonably small size of the mechanical arms. Although this method of attaining head tracking is very inexpensive, in comparison with other techniques, it also tends to be very unpopular due to the necessity of contact between the system and the operator.

An alternate form of sensing device are magnetic tracking systems which operate by utilising an emitter and receiver. The emitter sends out an
electromagnetic field which then induces currents in the receiver, known as eddy currents. The magnitude of the currents which are induced in the receiver determines its position and orientation. Although the operating principle is essentially the same there are two types of magnetic trackers (i) DC excited and (ii) AC excited trackers. The difference being in the technique used to generate the magnetic field.

In the presence of external magnetic fields or metallic objects, the position and orientation which will be generated by an AC tracking system is subject to errors, which can be likened to noise effects. The advantage with this type of system is the update rate is very fast and they have a very low latency of around 4ms (Polhemus 1992).

DC tracking systems are not susceptible to the presence of metallic objects within the magnetic field, but are however very effected by the presence of ferromagnetic materials. This is due to these materials distorting the magnetic field by reflection. DC magnetic tracking systems are commercially available and are reported to have update rates of around 100Hz, but unfortunately have latency times in excess of 17ms.

Although magnetic tracking systems are the most widely used and available form of tracking devices they are, as explained, very susceptible to the effects of magnetic field distortion and external magnetic fields such as those produce by CRT display screens. They also exhibit a great deal of advantages over other types of tracking devices. Magnetic devices do not experience 'line of sight' problems which are pronounced with acoustic and optical systems. The systems are very convenient to use as, both AC and DC systems, only require one transmitter and one receiver, the latter which are very light weight and so can easily be worn by operators. The
advantage of using an AC based magnetic tracker over a DC based system is due to the higher position update rate and the very low latency time. These systems are commercially available and are marketed by Polhemus, USA. This type of sensor is used with most of the higher capability virtual reality systems and a more detailed explanation of the functionality of the Polhemus system is given in section 4.4.1.

4.4.1 Operation of AC Magnetic Tracking System.

The AC excited magnetic tracking system marketed by Polhemus is known as Fastrack. Fastrack is a six degree of freedom electromagnetic sensor measuring three dimensional co-ordinate position, yaw, pitch and roll angles. The functional architecture of the system is shown in figure 4.2. It consists of a fixed magnetic-dipole transmitting source, a movable magnetic-dipole receiving sensor and associated driving electronics. Both the source and sensor consist of three mutually orthogonal loops or coils. Each loop of the source is in turn excited with an identical, both in phase and magnitude, driving signal which produces an excitation dipole.

The excitation of the source generates a magnetic field which in turn induces currents in the sensor. Each excitation of the source produces three linearly independent current vectors in the sensor. Thus for each cycle of three driving signals, one per axis, nine components are generated at the sensor.

Although there are nine component vectors available, there is sufficient information contained in the outputs arising from a single driving signal to determine the position and orientation of the sensor relative to the source.
Chapter 4: Mechatronic Design Considerations.

Figure 4.2 Functional Schematic of the Polhemus System.

The signals generated by the sensing coils are subjected to analogue to digital conversion and then processed by the dedicated computer processor. The processing involves the calculation of the position and orientation solutions and formatting according to the user specifications. The format of the output data and the system operation is definable by the user through issuing the Polhemus with ASCII (American Symbolic Instruction Code) commands over a RS-232 or IEEE488 link. The data which the system produces can be configured to be output as either ASCII strings or ANSI/IEEE standard floating point format. These values can be output via RS-232 serial or IEEE488 parallel links to a host machine.

The latency period of the Fastrack system is due to a number of factors. It consists of the time required to sample the magnetic fields, solve for the sensor co-ordinates, make the output available and to transmit this information to the host machine. The time taken to sample the magnetic fields, solve for the six co-ordinate values and configure the data into an output format is 5.5ms (Polhemus 1992). This data must then be
transmitted to the host machine. The time taken for transmission between the Fastrack and the host machine is very dependent on the format of the output and the connection which is used. For a RS232 connection the higher baud rate or transfer rate offers a lower transfer time. For example an ASCII data string containing all six degree of freedom values consists of 47 bytes (including status bytes) and with a transfer rate of 115.2Kbaud requires a transmission time of 4ms. Thus the combined latency time between motion of the sensor and the host receiving the output data is 9.5ms. The Fastrack system operates with an update of 120Hz, so every 8.3ms the position and orientation of the sensor are updated. A summary of the performance characteristics of the Fastrack is shown in table 4.3.

<table>
<thead>
<tr>
<th>Coverage Envelope</th>
<th>A hemisphere of up to 3.05m in diameter.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Accuracy</td>
<td>0.8mm RMS for X,Y,Z.</td>
</tr>
<tr>
<td></td>
<td>0.15° for orientation.</td>
</tr>
<tr>
<td>Resolution</td>
<td>0.0005cms/cm of range.</td>
</tr>
<tr>
<td></td>
<td>0.025° for orientation.</td>
</tr>
<tr>
<td>Latency</td>
<td>4ms from receiver measurement period to beginning of transfer from output port.</td>
</tr>
<tr>
<td>Update Rate</td>
<td>120 Hz</td>
</tr>
<tr>
<td>Output Format</td>
<td>RS232 up to 115.2Kbaud</td>
</tr>
<tr>
<td></td>
<td>IEEE-488 parallel</td>
</tr>
<tr>
<td></td>
<td>100KBytes/sec.</td>
</tr>
<tr>
<td></td>
<td>ASCII or ANSI binary format.</td>
</tr>
</tbody>
</table>

Table 4.3 Summary of Fastrack Performance.
4.5 Actuation System Considerations.

The proposed telepresence system has four degrees of freedom, each which must be capable of being individually and simultaneously moved and controlled. Essentially two types of actuator were suitable for use in the telepresence system, DC servo motors or stepper motors. It is preferred to utilise identical types of motor for all the degrees of freedom as this eases the design of the control system.

There are a number of disadvantages with the use of stepper motors within the mechatronic system. Stepper motors tend to be utilised in open loop type control, the accuracy and resolution of the system is then dependent on the motors not slipping or hopping counts. This is sometimes a problem with more low cost versions, but is however reported to be prevented by using 5-phase steppers which exhibit a higher stability and holding torque. The response and acceleration of stepper motors in comparison to DC servo motors is generally quite slow and the output is sometimes not smooth or linear. This could cause the camera system to exhibit a non-smooth motion profile which might be detrimental to the system performance.

Although more expensive than stepper motors, DC servo motors offer a more responsive solution to the actuation problem. Providing these actuators are used with a feedback sensor, such as optical encoders, they are quite easily subjected to closed loop control. DC servos often operate at much higher speeds than is actually required and so it is necessary to utilise some form of speed reduction gearing.

The type of speed reduction gearing which is employed must be given careful consideration as low quality systems can introduce non-linear
effects, such as backlash, into the system. These effects can degrade the performance of the system and also introduce difficulties with the system control. Through the use of higher specification gearboxes, such as Harmonic Drive systems, the effects of backlash within the system can be significantly reduced. This type of gearbox lends itself to the development of precision mechatronic systems.

4.5.1 Actuator Position Feedback.

To form a basic feedback loop it is necessary to measure the rotary position of the joint or actuator. This can be achieved in many different ways, but the most commonly used sensors in robotic and machine tool systems are optical encoders. The configuration of such a joint is shown in figure 4.3.

Figure 4.3 Joint Actuator/encoder Configuration.

Optical encoders provide position feedback of angular movement and are available in two forms, (i) absolute encoders which provide absolute position information and (ii) incremental encoders which provide incremental signals for recording the position displacement. Incremental encoders are used far more commonly than absolute encoders as they can
be manufactured with higher resolutions for a lower cost, however the advantage of absolute encoders is in the ability to directly move to a zero or reference position without the need for calibration or external referencing.

As the encoder rotates, digital pulses are generated and sent to the encoder interface. Incremental encoders generally have three output channels, channels A, B and C, as shown in figure 4.4.

The signals A and B produced by the encoder are out of phase by 90° and this phase difference provides a means of evaluating the direction of the underlying motion. For example when signal A is leading, signal B represents the rotation being in one direction and signal B leading signal A represents the reverse. Also with many rotary incremental encoders there is a further output channel C, which signifies the start or end of a revolution.

The resolution of incremental optical encoders is governed by the number of lines which the encoders have, which can range up to and above 1000 lines. The number of lines which the encoder has determines the number of pulses which would be produced by channels A or B, during a single complete revolution.

Both the signals A and B can be directly fed into up/down counters to measure the position displacement, with the counting direction (i.e. up or down) controlled by the phase difference between the two signals.

In most cases the incremental encoders are mounted directly on the output shaft of the motor. If a gear reduction unit is used the resolution is increased by a factor of the reduction ratio. Further increases in the
effective resolution of the encoder can be realised by forming what is known as a quadrature signal.

The quadrature signal, as shown in figure 4.4, is formed by utilising the phase difference between the two A and B channels to form a signal which demonstrates a count rate with a frequency of four times the A or B channel. Thus it is possible to increase the encoder resolution by a further factor of four. The quadrature signal is generally formed by an interface board which accepts signals A and B as inputs. The quadrature signal contains the position count and the direction of motion, and therefore whether to count up or down, is determined by a consideration of which input channel is leading.

---

**Figure 4.4 Incremental Encoder Signals.**
4.6 Control System Considerations.

A controller is required as the interface between the head tracking system and the servo positioning system, as shown in figure 4.5. The controller must provide the ability to perform initialisation of the system, system management, interpret the trajectory data and formulate this information into a suitable format to perform closed loop position control of the servo system. As part of the closed loop control system, it must perform the control algorithms and interrogate the encoders to determine the system position.

The performance requirements of the system often limits the type of processing architecture on which the controller is based. In this case the controller is required to do more than simply perform position control of the servo systems and thus the architecture required is slightly more complex.

The complete servo system is comprised of four degrees of freedom which must be capable of moving simultaneously and controlled independently in real-time. Hence the controller must be capable of performing independent and simultaneous control of the four degrees of freedom. The controller must be highly responsive so as not to introduce any appreciable latency into the system, for reasons explained in section 4.4.

Essentially four different systems were considered for the development of the controller (i) A commercially available control system, (ii) An in-house developed dedicated controller, (iii) Micro-controllers and (iv) A single board processor based controller.
Commercial based controllers, although offering a simple and easy to install controller they tend to be rather expensive. This type of controller usually consists of a graphical user interface (GUI) and a vast library of software functions. They exhibit great functionality, allowing the controller to be operated in a wide range of modes, such as velocity control or torque control. They also offer many different routines for trajectory profiles and path planning. These controllers do tend to be rather expensive and often require an additional host computer to configure the system. The functionality which a commercial system offers is far in excess of the requirements of the current generation of telepresence system and so the cost of purchasing all this additional capability can not be justified.

The development of an in-house dedicated controller would offer the most elegant of solutions as the controller could be designed specifically for the telepresence system. The performance could be optimised to be highly
responsive whilst still minimising the introduction of any latency into the system. Unfortunately, the design of such a system is not a trivial task and for an initial telepresence system it was thought to be unnecessary.

A more suitable method of controlling the system is through the use of micro-controllers. Micro-controllers are, in general, single chip processors with an abundance of input/output (I/O) capability. They generally have digital I/O, digital to analogue converters (DAC) and analogue to digital converters (ADC) present on the board. Micro-controllers are relatively cheap to purchase and are programmed in a variety of different ways, predominantly at a low level, usually in Assembler. The programs are stored in Electrically Programmable Read Only Memory (EPROM) and therefore any modifications to the programs requires the generation of new EPROM chips. Although offering a low cost solution to the control problem, micro-controllers do require programming in Assembler which is not a trivial task. In addition when writing the software it must be performed on a remote machine, compiled and then down-loaded. Each type of micro-controller requires its own compiler and if the software is tested prior to down loading then an emulator is also required on the remote machine. Due to the lack of experience in utilising such controllers and Assembler, it was decided that this would not be the most efficient method of achieving a solution.

Single Board Computers (SBC), as the name suggests, are computers which are contained on a single board. The SBC usually comprises of a central processing unit (CPU), various I/O capability such as serial and parallel ports, memory both RAM and ROM, hard and floppy disk controllers and a suitable back plane connection. The CPU can essentially be one of many types such as 68030 as used in a PUMA robot controller (Chen 1994) or a
PC based 386 processor as used in some satellites. The back plane connection determines the type of back plane bus to which the board can be connected again these range in specification but by far the most popular are the VME bus and the PC/AT bus. It should be noted that the bus type only governs the method by which the data is transferred around the computer system and so it is quite common to have a PC and VME bus version of the same processor board.

A back plane bus is required for SBCs as it is very rare that the processor board will perform everything required of the system, such as I/O and so a method of communicating with other local peripherals is required. This is achieved through the use of back plane bus and dependent on the type communication is; 8 bit for the PC bus, 16 bit for AT bus, or 32 bit for the VME bus. Essentially the choice of bus has very little effect on the system performance, but more importantly determines the availability and cost of additional hardware. PC/AT bus hardware tends to be much cheaper than the corresponding VME bus hardware.

Based on the relevant cost advantages and the availability of supporting hardware, the decision was made to base the controller around a PC bus architecture. This offers a low cost, easily expandable and easily implemented solution. Nowadays the most widely available processors for the PC bus are the 386 or 486 types of processor with clock speeds ranging from 25 to 66Hz. These processors are present in many desktop personal computers, but are also available in SBC format. The 486 processor has a much increased calculation speed and with the addition of a maths co-processor the speed is further increased.
4.7 Summary.

This chapter has developed the design considerations of a high performance experimental telepresence system, a summary of which is given in table 4.4. Although many of the design considerations were performed concurrently in practice, this chapter has approached them individually.

The initial problem was to evaluate the static and dynamic requirements of the system and to attempt to match the system performance to the human neck and eyes. This would ensure that the system is capable of achieving the dynamic performance required to achieve a convincing sensation of remote presence. A consideration was given to the modelling of the human neck motion and how this could be reduced and simplified to yield an acceptable model on which the system kinematic design could be based.

Each of the various sub-systems were then considered in detail and the requirements of each developed. Possible techniques or equipment which could be utilised to meet these requirements were then reviewed and suitable solutions determined.

Overall this chapter has drawn together many of the considerations, both in terms of the mechatronic and human factors aspects, which must be made when evaluating the design criteria of a telepresence system.
## Table 4.4 Summary of Design Criteria.

<table>
<thead>
<tr>
<th>Mechanical design.</th>
<th>4 degrees of freedom, incorporating pan, elevation and vergence.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degrees of freedom.</td>
<td>4 degrees of freedom, incorporating pan, elevation and vergence.</td>
</tr>
<tr>
<td>Weight.</td>
<td>Maximum of 10 Kg.</td>
</tr>
<tr>
<td>Space envelope.</td>
<td>Less then 300 X 300mm diameter.</td>
</tr>
<tr>
<td>Inter ocular Separation (mm).</td>
<td>60 - 200 mm, incrementally variable.</td>
</tr>
<tr>
<td>Range of motions (°)</td>
<td>Pan ± 60°, 400°/s, 2000°/s²</td>
</tr>
<tr>
<td>Velocity (°/s)</td>
<td>Elevation ±45°, 150°/s, 1000°/s²</td>
</tr>
<tr>
<td>Acceleration (°/s²)</td>
<td>Vergence ±45°, 400°/s, 4000°/s²</td>
</tr>
<tr>
<td>Camera system.</td>
<td>Composite PAL, S-VHS</td>
</tr>
<tr>
<td>Lens specification</td>
<td>Near point of 0.2m.</td>
</tr>
<tr>
<td>Display System</td>
<td></td>
</tr>
<tr>
<td>Immersive</td>
<td>Eyengen3 Head Mounted Display.</td>
</tr>
<tr>
<td>Non-Immersive</td>
<td>Tektronix. LCD shutter with circular polarising glasses.</td>
</tr>
<tr>
<td>Head Tracking Sensor</td>
<td>Polhemus Fastrack, AC electromagnetic 6 degree of freedom sensor</td>
</tr>
<tr>
<td>Data connection</td>
<td>RS232 serial, IEEE488 parallel.</td>
</tr>
<tr>
<td>Actuators.</td>
<td>DC servo motors with Harmonic Drive Gearboxes.</td>
</tr>
<tr>
<td>Position Feedback sensor.</td>
<td>Incremental Optical Encoders.</td>
</tr>
<tr>
<td>Controller Architecture</td>
<td>486 PC based processor, PC/AT bus.</td>
</tr>
<tr>
<td>Input/Output Required.</td>
<td>4 channel analogue output. RS232 serial, Encoder interface Digital I/O,</td>
</tr>
</tbody>
</table>
CHAPTER 5

SYSTEM DESIGN

5.1. Introduction.

The telepresence system is a classic example of a mechatronic system, incorporating actuation and microprocessor control of a mechanical system. Many of the design criteria have been established in Chapter 4 and suitable solutions for the head tracking, controller architecture and display system components identified. This chapter aims to bring together many of those criteria used to identify the requirements of the telepresence system and to present a view of the integrated telepresence system. It focuses on the mechanical design paying particular attention to the configuration of the elevation axis drive arrangement and the selection of the motors for all axes. The control system architecture is presented with a detailed explanation of the functionality of the controller and its associated components. Explanations are provided regarding the real-time techniques used to implement the control system and serial communications. Finally the control system and display system are presented as an integrated telepresence system.

5.2 Mechatronic System Design.

The mechanical system is to be driven by DC servo motors with Harmonic Drive gearboxes. Harmonic Drive gearboxes are high quality anti-backlash gearboxes which are often utilised in precision mechanisms.
In order to meet the performance criteria outlined in table 4.1, the performance of the DC servo motors must be suitably matched to the mechanical structure. This can be achieved by considering the mechanical inertia which the actuator is to drive and the dynamic characteristics of the actuator.

To initially select the actuator for a particular axis, the rotation speed required, as set out in table 4.1, is compared with the rated output speed of the actuator, taking into account the speed reduction produced by the Harmonic drive gearbox. The load torque which is required at this constant speed is then compared with the rated torque of the actuator. Through a consideration of the inertial loading on the actuator, the average torque which is experienced over trapezoidal velocity trajectory can be evaluated. The inertial loading on the actuator is determined by using the schematic drawing, Appendix A, to calculate estimates of the inertia of each axis. These detailed calculations are reported by Asbery (Asbery 1994) and yield the values shown in table 5.1.

<table>
<thead>
<tr>
<th>AXIS</th>
<th>PAN</th>
<th>ELEVATION</th>
<th>VERGENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inertial Loading</td>
<td>26 156Kgmm²</td>
<td>10 871Kgmm²</td>
<td>671.5Kgmm²</td>
</tr>
</tbody>
</table>

Table 5.1 Summary of Load Inertia, Referred to the Relevant Axis.

If the average load torque over such a cycle is less than the rated torque of the actuator then the actuator will meet the specification. Any deviation from this will usually cause the dynamic response to be reduced. An example of the motor selection procedure for the vergence axes (Asbery 1994) is given in Appendix B.
Using the data presented in Appendix B, it is possible to select DC servo motors which will meet the dynamic performance criteria of the telepresence system. The servo motors and gearboxes are supplied by Harmonic Drive and form a very compact and integrated servo-gearbox system, a summary of which is given in table 5.2.

The reflected inertia at the gearbox output shaft is the moment of inertia, referred to the output shaft, resulting from the sum of the motor inertia and the Harmonic Drive gear inertia. Through a comparison of the inertia values contained in table 5.1 and 5.2, it can be seen that the load inertia is much lower than the reflected motor inertia for the elevation and vergence axes. However, the load inertia of the pan axis is slightly higher than the reflected motor inertia. Consequently the pan axis motor response will not be optimum. It was necessary to slightly compromise the optimum dynamic performance such that the required output speed of 400°/sec could be achieved. Other motors utilising the Harmonic Drive type of gearbox do not offer the same output speed, being limited to 180°/sec. The motor selection was limited to a choice of Harmonic Drive systems, due to the preferential operating characteristics of harmonic type gear boxes.

Each of the DC servo motors are powered through linear power amplifiers. Linear power amplifiers receive an analogue voltage signal and subsequently amplify this signal to a current level which can drive the actuator. An input signal of ±10V will saturate the amplifiers, with an output voltage of ±24V.

Each of the power amplifiers operates from a 24V power supply and can supply currents up to 2A, so the combined system power supply should be at least a 24V DC, 8A supply. However, in practice the maximum current
which is drawn by the smaller vergence motors is far less than 2A (0.8A peak), so theoretically a lower performance power supply can be used. The power supply actually used to drive the amplifiers and servos is a 24V, 10A or 240 Watt DC supply.

<table>
<thead>
<tr>
<th>Motor</th>
<th>PAN</th>
<th>ELEVATION</th>
<th>VERGENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RH-14 6002</td>
<td>100 rpm</td>
<td>50 rpm</td>
<td>110 rpm</td>
</tr>
<tr>
<td>RH-8 3006</td>
<td>14 Nm</td>
<td>3.5 Nm</td>
<td>0.59 Nm</td>
</tr>
<tr>
<td>RH-5 5506</td>
<td>180 X 30 dia mm</td>
<td>132 X 33 dia mm</td>
<td>70 X 21 dia mm</td>
</tr>
<tr>
<td>Gearbox Ratio</td>
<td>50:1</td>
<td>100:1</td>
<td>80:1</td>
</tr>
<tr>
<td>Reflected Inertia at Gearbox Output</td>
<td>21 000 Kgmm²</td>
<td>15 000 Kgmm²</td>
<td>16 000 Kgmm²</td>
</tr>
<tr>
<td>Mechanical Time Constant</td>
<td>7msec</td>
<td>8.5 msec</td>
<td>13.3 msec</td>
</tr>
</tbody>
</table>

Table 5.2 DC Servo Specifications.

5.2.1 Mechanical System Design.

The mechanical system is required to exhibit the four degrees of freedom as established in figure 4.1. These are a pan motion, an elevation motion and the independent vergence motions of the two cameras. To minimise the space envelope and load inertia of the system, all the axes of motion are located as close as physically possible together. This also has the added advantage of minimising the separation between the axes of rotation, consequently reducing the relative inertial effects of each axis.
It is necessary to develop the vergence mechanism as individual modules such that the inter ocular separation can be easily re-configured. Additionally the mounting position of the vergence mechanism must allow the range of available separation to be easily increased/decreased if necessary. This type of requirement is very suited to a modular design philosophy which was adopted during the design phase.

From the initial schematic drawing, Appendix A, a three dimensional kinematic model was generated to evaluate the motion profiles of all the degrees of freedom. This was achieved using a three dimensional modelling and animation package, 3D Studio. An image of the model from 3D Studio is shown in figure 5.1.

![3D Studio Model of Telepresence Head](image)

Figure 5.1 3D Studio Model of Telepresence Head.

The three dimensional model of the head system enabled the kinematics to be viewed and evaluated prior to the manufacture of any part of the mechanical system.
As an additional demonstration of the capabilities of the system, the three dimensional model was imported into a virtual environment where it could be manipulated in real-time. The virtual world was generated using World Tool kit. This enables the system operation to be completely simulated. Through aligning the viewpoints of the virtual world with the camera positions and using the Polhemus head tracking system, a full simulation was created. Although a full simulation does not offer any significant advantages from a technical viewpoint, it does serve as a useful demonstration and marketing tool prior to the actual system build.

The mechanical system is manufactured from aluminium to minimise the weight of the system, to prevent corrosion of the aluminium surface the components were subsequently anodised.

Due to its modular design, the system essentially consists of two units, the pan and elevation system and the vergence system. The vergence attachment can be further reduced to its residual components, namely two vergence drive units and a ‘T’ bar mounting bracket. Each of the vergence drives mount onto the ‘T’ bar in the form of a saddle mechanism which is held in place by two cap screws. This allows the separation of the vergence mechanisms to be easily changed for experimental purposes. Separations in the range of 60 to 200mm in 10mm increments can be achieved using the current ‘T’ bar arrangement. The ‘T’ bar and vergence units are then easily attached to the pan and elevation unit by way of a further two cap screws.

All the outputs of the DC servo-gearbox units are used with a direct drive interface to the mechanical structure. The mechanical system can be driven directly by the DC servos without the need for flexible couplings as the mechanical components are manufactured to a very high tolerance to
prevent miss-alignment effects. A novel technique is utilised to form the direct drive mechanism of both the pan and elevation axes. At the output shaft end, the bearings present within the gearbox and servo are used to support rotation, as shown in figure 5.2.

This is possible due to the low radial loading which is experienced on the output shafts (Asbery 1994). At the opposite end to the output shaft, the structure is supported by bearings which locate around the motor body itself. In the case of the pan axis this takes the form of a pair of deep groove ball bearings which are mounted back to back and for the elevation axis take the form of a roller thrust bearing and a deep groove bearing.

![Diagram of Bearing Arrangement for Elevation Axis](image)

**Figure 5.2. Schematic of Bearing Arrangement for Elevation Axis.**

The system exhibits very smooth rotations on all axes with extremely little visible backlash. This is due to adopting direct drive of all axes from the outputs of the Harmonic Drive gearboxes, which are anti-backlash gearboxes.
The elevation axis is only capable of performing a limited range of motion. Any motion in excess of this range will cause one part of the mechanical structure to collide with another. A sudden impact loading of the gearbox, as occurs during a collision at high speed, could cause it significant damage. For this reason soft stops are positioned at the upper and lower extremities of the elevation motion which will be complemented by limits introduced in the controlling software. The actual mechatronic telepresence head is shown in figure 5.3
5.3 Control System Design.

The control system must perform three main tasks, the first is to initialise the system, the second is to format the trajectory data and the third is to subsequently drive the system to the desired position. The controller is based on PC/AT architecture and the control algorithms are performed in software. A simplified schematic of the control system configuration is shown in figure 5.4.

Developing the control algorithms in software enables the controller to be open architecture in design which simply means that the controller is easily configured by the user and readily interfaced to a variety of input sensing devices. Through programming the controller in software it would be possible to easily introduce additional algorithms or more advanced control schemes, such as predictive or non-linear control.

![Figure 5.4 A Schematic of Control System Configuration.](image)

The control scheme used consists of a three term or PID controller. Three term controllers consist of an integrator term ($K_i$), a differentiator term ($K_d$) and a proportional term ($K_p$). By tuning the three terms $K_i$, $K_d$, and...
K_p, the system dynamic characteristics can be modified, such that the settling time and overshoot are minimised, whilst still retaining an acceptable dynamic response.

A limited degree of modelling of the control system was performed using Matlab (Herskovitz 1989). The system was modelled as a linear system using a first order servo model, the pan axis model is as detailed in Appendix D, however the system model has a number of shortcomings. To achieve an accurate model of the system requires an account to be made of all the inherently non linear components and the time sampled controller and the inertia loading to be accurately known. Due to the mechanical structure being designed manually and not in a CAD system, the inertia values can only be approximated. Consequently the results of the system model were viewed with significant caution.

The model which is presented in Appendix D is just far to simple and ignores many parameters which will have an effect on the system performance. The model must include a consideration of the following effects; i) the time sampled controller, ii) the digital to analogue conversion effects, iii) friction effects in the bearings and gears, iv) an accurate evaluation of the load inertia's, v) higher order motor model and vi) non-linear components which are present such as the power amplifiers. Additionally for an accurate model of the elevation axis, it must include an account of the gravitational effects due to the elevation axis moving in a vertical plane.

In its most simplistic form the controller operates on the following principle. The trajectory data is generated by the Polhemus sensor and transferred to the controller via a serial communications link. The position
data is compared with the actual actuator position, by interrogating the encoder counters, and an error signal is generated which is then subjected to the PID control algorithm, the result of which is then passed to the digital to analogue converter. The digital level signal is then converted to an analogue voltage and output to the power amplifiers, which then acts as the input voltage to the DC servo motor. This cycle is then repeated at a constant rate, according to the controller sample time.

5.4 The Controller Architecture.

The controller is based on a PC/AT bus architecture using a commonly available PC processor. To function, the controller must consist of a processor, digital to analogue converter, serial communications and digital input/output (I/O).

Due to the ease of implementation, the controller was developed on a desktop PC, which eventually would become the dedicated telepresence controller. Of course this is by no means the most efficient use of such a machine, but the development is simplified and the time to completion far reduced. The processor used for the controller consisted of a 486SX-33Mhz processor which has a maths co-processor and a processor clock cycle of 33Mhz. In addition to the processor, the machine in which it is located has two serial interfaces, a monitor driver card and a 6-slot back plane.

Additional peripheral cards are required which consist of digital to analogue converter (DAC), encoder interface card and digital I/O. The operation of each of these cards will be explained and discussed in more detail.
As with installing any additional peripheral cards onto data buses or back planes, each card must be allocated a specific address. The address enables the various cards to be accessed through software, essentially specifying a location where a given card can be accessed.

5.4.1 Digital to Analogue Converter (DAC) Card.

The telepresence system consists of four independent degrees of freedom each with its own actuator. Therefore four digital to analogue converters are required. Nowadays, most DAC peripheral cards offer multiple channel outputs within a single card, hence only a single four channel DAC card is required.

Digital to analogue conversion involves converting digital levels to analogue output voltages. The resolution of the digital to analogue conversion process is dependent on the number of bits used by the DAC. For example a DAC with full scale output of ±10V, which utilises 8 bit conversion, will be capable of generating output voltages with a resolution of 0.08V, whereas a 12 bit converter is capable of 0.005V.

It is very common nowadays for DAC cards to be 12 bit, this offers significant resolution advantages over converters with a lower number of bits. A suitable 4 channel, 12 bit, DAC is manufactured by Amplicon Liveline and is known as the PC24. The PC24 offers a direct connection to the PC data bus and its operation is configured in software and by hardware jumpers. Hardware jumpers enable the board to be configured for either uni- or bi-polar output voltage ranges.
For the telepresence system, bi-polar operation is necessary to allow the actuators to rotate in both directions. The transfer function of the PC24 whilst operating in bi-polar configuration is demonstrated in figure 5.5.

![Figure 5.5 DAC Bi-polar Transfer Function.](image)

5.4.2 Encoder Interface Board.

The position feedback of the actuators is generated by pulse trains from incremental optical encoders. Each actuator has an individual encoder which is directly connected to the motor shaft and not to the output of the gearbox. This increases the effective resolution which can be realised by the encoders by a factor of the gearbox reduction ratio.

The form of the outputs of the incremental encoders is explained in chapter 4.5.1. The output consists of two pulse trains which are out of phase by 90°. To utilise this information as part of a feedback loop, the number of pulses generated by each channel must be counted. This is achieved by feeding the
pulse trains into an interface board, supplied by Keithly Data Acquisition. The interface board is located on the PC back plane bus and is allocated a location address, as with other peripheral cards on the bus.

The encoder interface board consists of four 24 bit counters, one per optical encoder. The pulse trains can be decoded to form quadrature inputs which has the effect of increasing the resolution of the given encoder by a factor of four. The resolution which can be achieved by a given encoder is given by equation 5.1.

\[
\text{Resolution} = \frac{360}{4 \times \text{Number of encoder lines} \times \text{Gearbox ratio}}
\]

The resolution of each axis is shown in table 5.3.

<table>
<thead>
<tr>
<th>Encoder Lines</th>
<th>Pan</th>
<th>Elevation</th>
<th>Vergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gearbox Ratio</td>
<td>50 : 1</td>
<td>100 : 1</td>
<td>80 : 1</td>
</tr>
<tr>
<td>Resolution</td>
<td>0.0036&quot;</td>
<td>0.0018&quot;</td>
<td>0.00313&quot;</td>
</tr>
</tbody>
</table>

Table 5.3 Resolution of Each Axis.

Each of the counters, when used in quadrature mode, is capable of counting A/B quadrature with a maximum input frequency of 0.33Mhz. As there are four counters on the board which may need to be latched and read at any one period in time, each one requires its own address. However, to save address space on the PC bus, the board utilises what is known as indirect addressing which allows the card to only occupy two direct ports on the bus. The first one as a command port, i.e. which counter is to be read, and the second for data flow.
The behaviour of the encoder interface board is configured both in hardware and software. All the counters can be latched and the data extracted, by sending the correct commands in software. Thus the controller software can interrogate the encoder positions by sending the appropriate commands to the board base address and then reading the values through the data port address.

5.4.2.1 Differential Driver Interface Board.

The inputs to the encoder interface board from the encoders can be in one of two forms differential or single ended. Differential signal usually means that the input consists of two wires, usually twisted a pair. The input signal is then derived by evaluating the difference in the signals from the two wires. This technique offers a high noise rejection ratio, as both of the channels will, in general, be subjected to the same noise and thus the difference in the two signal levels remains constant.

Single ended TTL signals are more commonly known as open collector signals. This simply means that the signal consists of a voltage on a single wire which is floating, hence must be pulled up to the supply voltage through a pull up resistor. Due to the open collector encoder signals, they are much more susceptible to the effects of noise. They are also limited in the distance over which the signals may be driven and in practice around 0.5m is usually used as a limiting value. For these two reasons it is preferable to use differential signals to transmit the encoder pulse signals.

Unfortunately, the encoders which are used for the vergence actuators, due to their compact size, are not available in differential output form. Consequently an interface board was designed to convert from open
collector to differential signals, for the two encoders used on the vergence axes.

5.4.3 Digital Input/Output Board.

The digital input/output (DI/O) board PET-48DIO, supplied by Industrial Computer Products (ICP), is also located on the PC back plane data bus. In addition to providing the capability of digital I/O, the board also has an on board clock source. The clock source, an Intel 82C53, can be programmed to generate timing signals which can be directed to form processor interrupt signals.

The DI/O board acts as an interface to a digital joystick, which can be used as an alternative form of demand input to the system. It also acts as an interface between limit switches on the telepresence system head and the controller. The board is configured in hardware by the connection of jumpers on the board and is capable of accepting 48 lines of DI/O, in the form of six 8 bit ports. The values of the I/O ports are accessed in a similar way to the encoder counters on the encoder interface board, through indirect addressing.

The board contains two clocks or counters which are both 16 bit counters that can be cascaded together to form a single 32 bit timer. The timer can be configured to operate in numerous different modes of operation by writing different control words to the control register. The most suitable mode for this application is in the form of a rate generator. In this mode one of the counters is set to clock a certain rate and then every, N, clock cycles the second counter will be clocked, as shown in figure 5.6.
As shown in figure 5.6, the output of the second counter can be directed to initiate a processor interrupt.

![Diagram showing operation of 82C53 in Rate Generator Mode]

Figure 5.6 Operation of 82C53 in Rate Generator Mode.

5.4.3.1 Interrupts.

Often computer systems are used to constantly poll a particular I/O device, this however means that whilst the processor is repeatedly carrying out the polling process it cannot do any other tasks. One possible way around this problem is to reduce the rate at which the device is polled. But if the device requires a rapid response from the processor as soon as its status changes, as is often the case with real-time systems such as controllers or instrumentation, then the delay in response becomes unacceptable. Any occurrence of this delay can be very much minimised by the use of interrupts. The telepresence control system is required to perform real-time simultaneous control of four degrees of freedom and so to perform continual polling of the trajectory I/O port would not be very efficient use of the
processor. The use of interrupts allows the amount of processor time spent checking for data to be reduced by only requesting that the trajectory I/O is serviced when new data has arrived. A digital control system requires the control process to be performed at a constant rate which is another suitable application for interrupt processing. Hardware can be configured to act as a rate generator and subsequently form interrupt signals at a constant time interval. This again allows the processor to be free from performing constant polling or monitoring of a timer.

In their simplest form, interrupts are signals which when activated cause the processor to stop whatever it is doing and do something else. Once this other task has been completed the processor will resume whatever it was doing prior to the occurrence of the interrupt.

So when an interrupt signal (IRQ) is activated the current flow of the program is suspended and the program branches out to another routine, known as the Interrupt Service Routine (ISR). The ISR is performed and then an end of ISR signal is activated, when the main program will regain control and continue as normal.

Interrupts can be generated from both software and hardware devices. Hardware interrupts occur in response to electrical signals received from a peripheral device such as an I/O board, serial communications port or hard disk controller.

The PC/AT architecture can manage 16 IRQ levels and it is possible for more than one interrupt to occur at a single period in time and so a device is required to control and keep track of the IRQs which require attention. In a 386/486 PC based system this is known as the Programmable
Interrupt Controller (PIC) and is generally the Intel 8259A chip. The 8259A allows any one interrupt to be fed into the processor and also prioritises the interrupts such that a lower level of interrupt does not interrupt the processor whilst it is servicing a higher priority interrupt.

Interrupts and ISRs are easily implemented in software. The specific IRQ number is allocated a vector address which corresponds to the location of the ISR in RAM memory. This is generally in the form of a far pointer, due to the location which is designated for ISRs. Thus when the IRQ is activated the program will jump to the location of the ISR and then after completion will return to the main program.

### 5.4.4 Serial Communications.

The Polhemus Fastrack six degree of freedom sensor is connected to the controller by means of a RS-232 serial communications link which is a very common technique for the transfer of data between remote devices and also local devices such as the mouse used on a desktop PC. It is capable of supporting hand-shaking in the communication process, but the Polhemus system operates in null modem form which is without hand shaking.

To operate in null-modem form three connections must be made between the controller and the Polhemus sensing device. These include connecting the transmit pin, TX, of the controller to the receive pin, known as RX, of the Polhemus, connecting the receive pin of the controller to the transmit pin of the Polhemus and a connection between the logical grounds of the two ports.
The interface between the RS-232 port and the PC processor is realised by a chip-set which is manufactured by National Semiconductors. The most important chip of which is the NS8250 Universal Asynchronous Receiver Transmitter (UART). The UART can be programmed in software to configure its operation. Some of the options which can be configured include the baud rate for data transfer, start and stop bits, parity bits and interrupt enables.

The baud rate determines the number of bits which are transmitted per second and for this particular application where short time delays are paramount, this has been set at the maximum rate of 115.2 KBaup. The baud rate only governs the rate at which bits within a byte are transmitted and not the rate of byte transmission. The stop and start bits signify to the UART that a byte of data is beginning or ending and in this system the data transmission is configured to utilise only one stop bit per data byte. The parity bit is an additional bit which may be introduced into the data to generate even or odd parity depending on the setting. This is used to error check the data, as a byte exhibiting even parity must exhibit an even number of logical highs otherwise there was an error in the transmission of data.

Although the data generated by the Polhemus tracking system is supposed to be transmitted at a fixed rate, it was thought that this rate was not particularly exact and could not be relied upon. This problem was easily overcome by the introduction of real-time communications achieved by programming the 8250 UART to generate an interrupt signal when data arrives at the serial port. As previously explained in section 5.4.3.1, the initiation of an interrupt request causes the main program to be suspended while the interrupt request (IRQ) is serviced by the interrupt service
routine (ISR). In this case the ISR involves the reading of the data which is present at the serial port and storing it in a trajectory buffer.

5.5 Implementation of the Controller.

The controller is implemented in software on a 486SX 33Mhz PC, with the DAC, encoder interface and DI/O boards installed on the PC bus. The functionality of the control system is shown in figure 5.7.

The timer/counter which is present on the DI/O board is programmed to generate an interrupt signal every 2ms. This time period is used as the controller sample time and the interrupt which occurs causes the PID control algorithm to be performed, as an ISR.

![Figure 5.7 A Functional Schematic of the System Controller.](image)

PhD. 1997 Richard Asbery
The initial steps performed by the controller involve the initialisation of the various boards, the Polhemus sensor and carrying out a self calibration routine. The Polhemus sensor is configured to continuously supply only the roll, elevation and pan data in the IEEE floating point format which eliminates the overhead of transmitting position data which is not directly utilised by the controller.

Once the peripheral boards and the Polhemus have been initialised the software enters into tracking operation. This takes the form of a continuous loop until an interrupt arrives from either the serial port or the 2ms clock. The serial port interrupt signifies that there is a new trajectory data set available and this data is then collected from the serial port and stored in a trajectory buffer.

Each time the DI/O timer generates an interrupt, which occurs every 2ms, the trajectory buffer is interrogated. This interrogation only occurs if the contents of the data buffer has changed since the last timer interrupt. The trajectory data is then converted to the corresponding number of encoder counts and the PID algorithm applied to the resultant data. A significant amount of error checking is performed before the output of the PID algorithm is loaded to the DACs via the PC bus. This ensures that the demand trajectory lies within angular limits configured in the software. These limits are currently +90° for the pan axis, +45° for the vergence axes and +60°, -45° for the elevation axis.

Once the control output signals have been limit checked, the values are sent to the DACs. The DACs can only be loaded with hexadecimal values in the range of 0h to FFFh, where 0h and FFFh correspond to positive and negative full scale outputs respectively. The control values are then
correspondingly conditioned to account for the DAC configuration and saturation characteristics. The conditioned values are then loaded into the DACs and subsequently the analogue signals are passed to the power amplifiers and hence the actuators are driven to the demand position.

This whole process occurs repeatedly as the control cycle and operates on a sample time of 2ms, which again can be easily re-configured if necessary. Figure 5.8 demonstrates the system undergoing head tracking control. The operator is wearing a head mounted display (HMD) which displays the images collected by the cameras.

![Figure 5.8 System Undergoing Head Tracking.](image-url)
5.5.1 Self Calibration.

In its powered down state the telepresence head can be resting in a number of varied orientations. This can cause problems during the initialisation process, as there exists some ambiguity as to the location of the various joints. It is preferable for the system to automatically wake up and position itself in the \textit{look straight forward} orientation. One strategy for achieving this is through the use of self calibration limit switches.

The use of limit switches allows each joint to be driven to an initial position, the angular location of which is known relative to the zero position. The limit switches are connected to the DI/O board, which when activated cause a corresponding logic change on the board. Once the limit switch has been tripped the joint is driven to its \textit{"look straight forward"} position. This has been implemented for the elevation axis and could be very easily introduced to the pan axis. So, from whatever position the system is powered up from, the elevation axis will automatically resume the zero position from where the control loop will resume control of the system.

5.6 The Display System Implementation.

The display system provides the visual stimulus to the operator of the telepresence system. The images are collected through the use of two Panasonic WV-KS152 colour CCD cameras. These cameras are of the remote control unit type, and exhibit a very small space envelope and mass at the actual image pick up end. The space envelope is around 40 x 12mm diameter with a mass of about 16g. The camera control units are genlocked.
together and are connected directly to the head mounted display and stereo switch, as shown in figure 5.9.

A left/right frame synchronisation signal is generated by the Scan converter, which also acts as a frequency doubler. In addition to forming a left/right field sequential signal the stereo switch acts as a master genlock signal generator, thus both of the cameras and the stereo switch are synchronised together. This ensures a synchronised camera and display configuration.

![Diagram of Display System Configuration](image)

**Figure 5.9. Schematic of Display System Configuration.**
5.7 Summary.

The telepresence system utilises a controller based on a 486SX 33Mhz PC to interpret the orientation data generated by an electromagnetic six degree of freedom sensor. The controller, which operates on a control cycle of 2ms, subsequently drives four DC servo motors independently and simultaneously to the demanded positions under closed loop control. The controller has been developed in software to enable easy reconfiguration or the introduction of more advanced control strategies such as predictive tracking. The control system has been designed such that it incorporates real-time programming techniques to maximise the performance of the control system on the architecture utilised. It also allows the simple interface of additional input devices via a serial port or digital input.

The mechanical design of the telepresence head incorporates four degrees of freedom, which have been designed such that the dynamics simulate those of a human operator. The inter ocular separation of the two cameras is easily modified to allow future experimentation.

The output from the dual CCD cameras can be displayed simultaneously on a Tektronix stereoscopic display and on a Eyegen3 head mounted display (HMD). The HMD allows the operator to gain a sensation of immersion in the remote environment observed by the cameras.
CHAPTER 6

EVALUATION OF THE TELEPRESENCE SYSTEM PERFORMANCE

6.1 Introduction.

There are a number of different areas within which the performance of the telepresence system can be evaluated both in system and operator oriented terms. The operator oriented evaluation involves assessing how the system performs whilst being used during a remote operation task and the system oriented evaluation includes aspects of the mechatronic system performance such as controller step response, overall system latency and the optical characteristics of the system. This chapter focuses on the evaluation of the mechatronic performance of the telepresence system.

The determination of the system latency is of particular importance as it is a key attribute of a telepresence system. If the latency period is too high then the system becomes very difficult to use and a very small increase in the latency period can cause the sensation of remote presence to be lost. Analysing the step response of the controller is a fundamental aspect of controller design and in this case is very important, a system which drastically overshot and then oscillated about the demand position would prove very difficult to use. Therefore, the system should exhibit a minimal overshoot and a very short settling time to ensure that it is comfortable for use.
The optical characteristics of the system are also investigated and an assessment of the stereoscopic magnification is presented.

### 6.2 Step Response.

The response of a system to a step input request enables the dynamic characteristics of the output, which in this case is angular position, to be evaluated. A servo position control system is inherently a second order type of system and characteristics such as the rise time, overshoot and settling time can be used as a measure of the dynamic performance. It can also demonstrate whether an optimal solution has been achieved for the tuning of the PID control parameters.

![Step Response Diagram](image)

**Figure 6.1 Pan Axis Step Response.**

The step response of the system is tested by giving a 45° step input to each axis in turn and observing the corresponding encoder output during the motion. The response of the pan axis to a step demand signal is shown in figure 6.1. The figure shows two traces; one for the uncompensated system...
with $K_p=1$, $K_i=K_d=0$, figure 6.1a, and the second for the compensated system, figure 6.1b. In the compensated system case the PID parameters which were determined by experimental investigation were set to be $K_p=0.4$, $K_i=0.5$ and $K_d=0.005$.

The response of the controller to a step input response is shown to be very fast, reaching the set point in a time of 0.15 seconds. If a linear motion profile is assumed during the rise time, then the velocity of the joint can be approximated. Through using equations describing simple linear motion, the joint velocity can be evaluated to be approximately $300^\circ$/sec.

The response of the elevation axis to a $45^\circ$ step input is shown in figure 6.2. Again the figure shows two traces, one for the uncompensated system, figure 6.2a, where $K_p=1$ and one for the compensated system, figure 6.2b, where $K_p=0.3$, $K_i=0.01$ and $K_d=0.001$. Again the PID parameters were determined by experimental investigation.

![Figure 6.2 Elevation Axis Step Response.](image)
The two traces shown in figure 6.2, exhibit limited differences. The main differences are that the settling time and the initial overshoot has been reduced for the compensated case. Again, through assuming a linear motion profile the joint velocity can be estimated. The elevation joint reaches the set value, $45^\circ$, in 0.2 seconds which is equivalent to a joint velocity of $225^\circ$/sec.

In both cases, an empirical search approach has been adopted to determine the PID controller parameters. This technique offers a very quick method to determine reasonable controller parameters, as shown by the compensated plots of figures 6.1b and 6.2b. However, this is a less than ideal solution and it would be expected that a further increase in the system response performance could be achieved by using parameter optimisation techniques which are freely available in software form or by further experimental searching.

### 6.3 System Latency.

The latency or delay of a telepresence system is of paramount importance when considering the performance of such a system. The latency of a telepresence system is defined as the time interval between the initiation of the sensor motion and the mechanical system responding to this sensor motion.

Many functions of the system introduce a time delay and the sum of all these contributions is equal to the total system latency. The most notable, and most commonly investigated, of these contributions is the effect of sensor lag (Adelstein, Johnstone and Ellis 1992). The sensor lag is the time taken from the onset of head motion to the time when this measurement is available at
the controller interface. This includes the time taken for the transmission of the co-ordinate data over the serial communications line.

Other sources of time delay are the communication transmission time between the controller and the remote camera system, the calculation period of the controller and the inertial effects of the mechanical system. The controller time delay is generated during the process of decoding the trajectory data and digital to analogue conversion. The inertia of the mechanical system affects the magnitude of the latency period due to it acting as a resistance to the initiation of motion.

An attempt has been made to measure the overall system latency, between the initial movement of the Polhemus sensor and the corresponding representation of that movement by the mechanical system.

6.3.1 Technique Adopted to Evaluate the Latency.

The latency of the system can be separated into two distinct areas; (i) the latency developed by the Polhemus sensor and (ii) the latency generated between the time the trajectory data arrives at the controller and the time the mechanical system begins to respond. Approaching the problem in these two distinct areas allows the mechatronic system latency to be specified independently to that of the particular sensor being used.

To evaluate the system tracking latency an operator was asked to generate two different types of head motion trajectory for the pan axis, one which involved a relatively slow position change at a constant velocity and a second trajectory which was far more dynamic and which included changes in direction, similar to a sinusoidal input. The same operator also generated a
dynamic trajectory for the elevation axis. Two types of trajectory were used in order to investigate if variations in demands placed on the mechatronic system made any significant effect on the magnitude of the latency period.

During testing, two sets of data were recorded, one of which showed the demand trajectory produced by the Polhemus and the other which showed the corresponding position of the servo system. Each time the control cycle was performed (every 2ms) the trajectory pair were recorded.

In order to simplify the method for evaluation of the system latency, the trajectory data which was produced by the Polhemus sensor was only recorded when it arrived at the controller. Thus this data has already been subjected to a certain degree of time delay. This time delay is composed of the serial communication transmission time and the inherent delay which is introduced by the sensor itself. However, this does not introduce any significant errors into the latency evaluation, due to their magnitude being easily calculated theoretically. The sensor time delay is well documented (Adelstein, Johnstone and Ellis 1992) and the serial communication occurs at a standard rate which in this case is 115.2Kbaud. The Polhemus latency period is in the order of 5.5ms (Polhemus 1992) and for the transfer of a complete data set, using a 115.2Kbaud serial link, is 4ms. The combined latency period, including the effects of the 8.3ms update rate of the Polhemus is 17.8ms.

6.3.2 System Tracking Results.

The two operator generated trajectories for panning motions are shown in figures 6.3 and 6.4. A similar dynamic trajectory for the elevation axis is shown in figure 6.5. These figures show the sensor trajectory and the corresponding servo system trajectory. The sensor trajectory does not take
into account the delay which is introduced between the actual motion of the operator’s head and the serial transmission delay. If these effects were included it would cause the servo trajectory to be further shifted to the right, along the time axis relative to the Polhemus trajectory, by 17.8ms.

The data contained in the following figures appears to be very coarse for the sensor trajectory. This is due to the data being sampled every 2ms, when the Polhemus sensing system only updates its position once every eight control cycles or approximately 16ms.

Figures 6.3, 6.4 and 6.5 clearly show the servo system position lagging behind the position of the Polhemus sensor. At any given time the position of the sensor can be seen to be in front of the position of the servo, the magnitude of which is shown as the error curve.

![Figure 6.3 Pan Axis Approximate Constant Velocity Trajectory Tracking.](image)

Figure 6.3 Pan Axis Approximate Constant Velocity Trajectory Tracking.
To establish the actual latency time the servo time history must be shifted along the time axis such that the error between the sensor and servo trajectory is minimised. This has been performed for both the simple constant velocity trajectory and the more complex dynamic trajectory, plots...
of which are shown in figures 6.6 and 6.7 respectively. It has also been performed for the elevation axis and as shown in figure 6.8.

Figure 6.6. Pan Axis Time Shifted (60ms) Trajectory.

Figure 6.7. Pan Axis Time Shifted (60ms) Trajectory.
In order to minimise the position error between the sensor and servo trajectories, the servo trajectory must be shifted in time by 60ms. Thus the time lag between the Polhemus data arriving at the controller and the servo system realising this trajectory is 60ms. This shows that the system latency time is the same for both the pan and elevation axes. From a consideration of figures 6.6, 6.7 and 6.8, it appears that the latency is not velocity or trajectory dependent, but is a function of the system transport and calculation delay.

If an account is made of the Polhemus sample, calculation and transmission time then the total system latency is in the order of 78ms (Asbery and Pretlove 1995). This latency time is less than of 50% of the time quoted by the NASA system (Bolas and Fisher 1990). The low latency time equates to an improved operator sensation of presence when the interaction of the operator and the remote camera system is considered.
The vergence axes are not subject to dynamic sensor input, hence no tracking plots are shown.

6.4 Stereoscopic Magnification.

The concept of stereoscopic magnification, $M_s$, for a dual camera system was developed in Chapter 3. This generated equation 6.1, which defines stereoscopic depth.

$$M_s = \lim_{\Delta \to 0} \frac{2 f_c L \tan \phi}{\Delta (L-f_c)}$$

where $f_c$ is the focal length of the camera lenses, $L$ is the distance between the convergence point and the front nodal points of the lenses, and $\phi$ is the vergence angle. Stereoscopic magnification is important as it clearly demonstrates the effect on the stereoscopic image of variations in the interocular separation of the cameras.

Equation 6.1 is in the form of a limit, the limiting value of which is the separation between two objects, on to one of which the cameras are converged. The limit can be evaluated for a given camera configuration, Appendix C, to form a set of curves relating the stereoscopic magnification to the interocular separation, as shown in figure 6.9. It clearly demonstrates that for a lens focal length of 7.5mm, the stereoscopic effect is made significantly greater by increases in the interocular separation of the two cameras. The curves also demonstrate that the stereoscopic acuity tends towards zero as the camera to object distance tends towards 6m.
For an interocular separation of 6.5 cm, which is the average human eye separation (Hofsetter 1972), the stereoscopic effect is shown to be significant at object viewing distances of less than 2 m. At distances greater than 2 m the effect is less pronounced. At a viewpoint distance of greater than 3 m there is little difference in the stereoscopic effect which can be achieved by trebling the interocular separation. Consequently there would be little advantage in utilizing a larger interocular separation for viewing objects at this range.

6.5 Summary.

These experiments have evaluated the mechatronic performance of the telepresence system. An empirical approach to the evaluation of the PID
control parameters has enabled a reasonable step response of the system to be achieved. The step response of the system has shown to be such that it exhibits minimal overshoot of less than 1° in the elevation and pan axes and a low settling time. This performance is acceptable to an operator as it generates very little observable oscillation due to the speed at which the joints rotate. However, it is still anticipated that this performance could be further improved by optimisation of the PID control parameters or by additional experimental searching.

The investigation of the system latency period has shown it to be in the order of 78ms. This has been determined by considering the latency of mechatronic telepresence system separately from the sensor system. The sensor system was excluded from the assessment of the latency period as this would require further equipment to impart a controlled motion on the sensor, such that its actual position at any period in time could be determined. This would significantly increase the complexity of the experiment and would necessitate the introduction of other transducers which would most likely also exhibit a delay in their measurement cycle. Through only using the features of the telepresence system to establish the latency period, it has been determined very simply by analysing the motion profiles generated by the Polhemus sensor at the interface with the controller and the motion measured by the motor encoders.

The latency period exhibited by the telepresence system when combined with that of the Polhemus sensor system is significantly lower than other current telepresence systems which can be attributed to the integrated mechatronic design and real-time control techniques implemented in the control system. The aim of telepresence is to provide a seamless link to the remote system, such that the operator cannot identify the existence of the remote system.
In order for this to be achieved requires a low latency period, as if there is a delay between the motion of the operator's head and the change in scene which is presented to the operator via the HMD then the sensation of presence is immediately subdued. Consequently, an ideal system would exhibit a zero latency period but as this is not possible a minimal latency is preferred. Further experiments will subsequently be reported which investigate the effect of the magnitude of the latency period on the operator performance and comfort.
CHAPTER 7

TASK PERFORMANCE EVALUATION

7.1 Introduction.

The area of task oriented performance evaluation involves the application of the system to a number of experimental tasks such as a pick and place exercise and target trajectory tracking. The pick and place task allows a comparison to be demonstrated between the use of stereoscopic and monoscopic display systems. A simple experiment is performed to demonstrate the stereo advantage by using the telepresence system as a visual interface for a telerobotic cell. This investigation supports many other experiments which have been performed in this field.

A further group of experiments which also assess the operator oriented performance of the system are in relation to the type of the input device and the magnitude of the latency period. These investigations take the form of a number of target tracking experiments. The experiments allow the advantages of using a head tracked telepresence system over a conventional joystick controlled system to be evaluated. Also, using the target tracking experiment, the effect of increasing the total system latency period is investigated in order to establish and demonstrate a correlation between an increased latency period, degraded operator performance and operator comfort. The target tracking experiments require the formulation of a number of mathematical relationships between the target co-ordinate frame and the telepresence co-ordinate frame. A detailed explanation of the theoretical and practical development of this relationship is presented.
A comparative assessment of the display device is also performed. The aim of this experiment is to investigate the differences in the capability of an operator to perform a remote task whilst using two different types of video display system. The experiment utilises both immersive and non-immersive display systems to perform identical target tracking tasks with the same user input device.

This chapter presents the operator oriented experiments used to assess the performance of the telepresence system when applied to a number of different tasks. For each type of experiment the results and a detailed discussion are presented.

### 7.3 Demonstration of Stereo Advantage.

These experiments aim to demonstrate the enhanced operator performance which can be achieved through the use of a head slaved stereoscopic telepresence system. Two types of experiment were performed; the first was a simple pointing task and the second was a more complex navigation task.

Both of these experiments were performed using a Unimation Puma 560 mark III industrial robot and an open architecture robot controller (Chen 1994). Teleoperation of the robot is achieved through a keyboard interface between the controller and the operator which allows the operator to control the Cartesian position of the end effector only (i.e. three degrees of freedom). It enables simple control to be achieved using the numeric pad of a standard keyboard.

The target arrangement which was used in both experiments consists of a board on which pegs of differing heights are located. The target board is of
dimensions 600mm X 600mm and is configured with 20 randomly placed pegs with heights in the range of 50 to 150 mm. The workspace of the robot which contains the target board is viewed by the telepresence system, as shown in figure 7.1.

![Figure 7.1 Teleoperation Work Cell.](image)

The experiments were intended to demonstrate the increased operator performance in terms of the time to complete a given task. In both experiments the tasks were performed in three operational modes. The first method was to use direct viewing of the work space, the second case was to use the telepresence system in stereoscopic mode and finally, the third case was in monoscopic mode.

This allowed a comparative study to be made between the three approaches. For both experiments where the telepresence system was used, identical cameras were used for both stereo and monoscopic viewing. This ensured that any effects of differing optical arrangements, such as the field of view,
were minimised. For these experiments the cameras were arranged with an interocular separation of 70mm.

Due to the limited field of view of the telepresence system, it is not possible to observe the whole workspace without generating motion of the camera system. To view the complete workspace and to track the end effector of the robot, the operator must appropriately control the gaze of the camera system.

### 7.2.1 Pointing and Locating Experiment.

Using the target board placed in the teleoperation cell, the operator is requested to drive the robot pointer to touch a particular numbered peg selected at random and the time required to complete the task recorded. This process is repeated for five randomly selected pegs all of which have different heights and spatial locations. Each peg selected is known as a task numbered from 1 to 5 and the time taken for the operator to position the pointer as close as they felt possible to the top of the chosen peg is recorded. At this position the horizontal and vertical distance between the top of the peg and the pointer is measured using a rule and the RMS error calculated.

The experiment was repeated for each of the three different viewing configurations, (i) stereoscopic, (ii) monoscopic and (iii) direct viewing. In order to allow a direct comparison of the performance using the different viewing configurations the same pegs were used for all three sets of experiments. To prevent any biasing occurring in the experiment a sample of five operators were used to perform the experiment. All of the operators had no experience of using the teleoperator and were therefore given a five
minute training session in which to become familiar with the controls and the telepresence system.

The telepresence system required the operator to directly control the gaze direction through the motion of their own head. In order to operate in head tracked monoscopic mode, the output from the right camera was fed in to both of the displays within the Helmet Mounted Display (HMD). This ensured that a similar field of view and resolution was used for both of the remote viewing approaches. In the case of direct viewing, the operator was seated such that their eye position corresponded to the location of the remote camera system.

7.2.1.1 Results of Pointing and Locating Experiments.

The results of the pointing and locating experiment are shown in figures 7.2 and 7.3. The task number corresponds to the operator being asked to move the manipulator as close to the same numbered peg on the target board. Figure 7.2 clearly shows that the direct viewing configuration offers the fastest time to completion of the pointing tasks. The general trend is for the stereoscopic case to be completed faster than the monoscopic case. Additionally if the accuracy to which the task was completed is considered, as shown in figure 7.3, the stereoscopic case can clearly be seen to offer a much lower position error.
Chapter 7: Task Performance Evaluation.

Figure 7.2 Time to Completion of Task.

Figure 7.3 Mean Horizontal Positioning Error.

7.2.2 Wire Tracking Experiment.

The second part of these experiments involved using the same five operators to perform a different type of task which involved navigating and tracking the manipulator through a maze. The maze is formed from the same target board which was utilised in the pointing and locating experiment.
One end of a piece of wire was attached to the target board at point A, and the other end attached to the end effector of the manipulator. The aim of the experiment was to wind the wire around pegs B, C and D in turn. This requires the manipulator to be guided through the array of pegs whilst being tracked by the telepresence system. The time taken to complete the task and the number of collisions encountered were recorded. Again this experiment was performed in each of the three observational modes, (i) direct viewing, (ii) monoscopic viewing and (iii) stereoscopic viewing and an example of a completed wire tracking task is shown in figure 7.4.

Figure 7.4 Completed Wire Tracking Task.

7.2.2.1 Results of Wire Tracking Experiment.

The average time to completion of the wire tracking task by each of the five operators is shown in figure 7.5. This figure shows that the general trend is
for the stereoscopic observed case to be performed quicker than the monoscopic case.

The average number of collisions between the robot end effector and the target board was 0.8 and 0.4 for the monoscopic and the stereoscopic case respectively. A value of less than unity denoting that collisions did not occur with every operator.

![Bar chart showing time to completion for different operators and viewing conditions](image)

**Figure 7.5 Time to Completion of Wire Tracking Task.**

### 7.2.3 Discussion of the Pointing and Locating and Wire Tracking Experimental Results.

The experiments show that for a simple pointing and locating task the shortest average time to completion is achieved through the operator directly viewing the workspace without any optical aids. Stereoscopic viewing offers the second fastest time to completion, followed by the monoscopic case. It can be noted from figure 7.2, that in some instances the stereoscopic and monoscopic times to completion for a given task are very similar. However, if figure 7.2 is used in conjunction with figure 7.3, it can be
seen that although the completion times are in some cases similar, the positioning error is significantly increased with the monoscopic viewing system.

During the wire tracking task the operators experienced greater confidence in the location of the robot end effector relative to the pegs on the target board when operating in stereoscopic mode. This was demonstrated with a lower mean number of collisions for the stereoscopic case.

The increase in operator confidence with respect to spatial localisation of the target board was also demonstrated by the techniques used to move the end effector to the targets during the pointing tasks. The operators tended to adopt different search techniques to position the pointer at what they viewed as a suitable location when using stereoscopic and monoscopic viewing configurations.

Whilst using the monoscopic viewing configuration the operators tended to perform a great deal of searching along the gaze direction of the camera system. This was an attempt to determine whether the pointer on the robot was in front or behind the target. Whereas, in the stereoscopic observation mode the operator tends to move the robot directly to the target due to the enhanced precept of depth which is achieved.

During the performance of both the pointing task and the tracking task the operator was continuously visually searching the workspace to assess the difference between the current position and the required position of the robot. It was only when the robot became very close to the target that the operators head and hence the telepresence system remained stationary.
Throughout all the experiments the mean positioning errors are larger than
would be expected and for the direct viewing case the error would be expected
to be zero. This shortcoming can be attributed to the less than ideal
mechanism used to control the position of the robot. The keyboard interface
generated a number of difficulties. It was very easy for the operator to lose
the positioning of their fingers over the keys which controlled the robot
position. This was particularly evident in the stereoscopic and monoscopic
experiments where the Head Mounted Display was utilised and the operator
could not actually see the position of their own hands.

Due to the data buffer which is inherent in the keyboard design, at times the
robot would quickly move in an undesired direction, thus startling the
operator and appearing to be out of control. Although this is less than an
ideal technique for controlling a teleoperator, it has enabled preliminary
results to demonstrate the superior performance which can be achieved
through the use of a stereoscopic viewing system. The comparison between
the three viewing cases is comparable with that produced in other work in
the field (Cole and Parker 1989, Drasic, Milgram and Grodski 1989, Miller
and Mitchell 1990).

7.3 Target Trajectory Tracking Experiments.

These experiments aim to evaluate the operator performance achieved
whilst using the telepresence system to track the trajectory of a randomly
moving target which is similar to the process which would be adopted to
track and intercept a moving object. The experiments will allow a
performance comparison to be made between a range of three different input
devices to the telepresence control system, each one exhibiting distinct
characteristics.
The input devices comprise of (i) A Head controlled Polhemus system, (ii) A digital joystick and (iii) A Polhemus based joystick. In order to minimise the experimental variables and generate a consistent set of results, an identical display system is used for each of the three input devices. The display system used in these experiments is an Eyegen3 display which is an immersive head mounted display.

7.3.1 Experimental Configuration.

The aim of the experiment is to simultaneously record the location of a target and the direction of view of the telepresence system when directed towards the same target, thus allowing a measure of the tracking accuracy to be evaluated. This requires the target to be moved on a trajectory of which the position in three dimensional space can be determined at any given point in time. A simple approach to generating a controlled target trajectory is to mount the target on the end effector of a robot which is then programmed to move the target in a pseudo random trajectory. A pseudo random trajectory is required in order to present a target trajectory which is not easily predicted. For example, if a parabolic target motion profile was used the operator would find it very easy to guess the future position of the target. Therefore, this would not demonstrate the target tracking capability of the system, hence using a pseudo random target trajectory the effects of trajectory prediction can be significantly reduced. Also a pseudo random target motion is much easier to implement on the robot controller than a purely random trajectory. For these experiments a Unimation Puma 560 Mark III industrial robot with an additional open architecture controller (Chen 1994) is used to generate and monitor the target motion.
There are two main features which are required in order to perform the
target trajectory tracking experiments. The first is to synchronise the
sampling of the target position and the viewpoint position. The second is to
determine the location of the viewpoint which corresponds to the position in
space where the operator is looking. The target location can be determined in
the robot co-ordinate system by simply interrogating the robot controller to
generate the robot X, Y and Z co-ordinates and the orientation of the direction
of view can be established through interrogation of the telepresence
controller. It is however, the synchronous sampling of this position and
orientation measurement which must be accomplished.

The position and orientation sampling of the target position and the direction
of view can be performed synchronously by interfacing the two separate
control systems. This requires forming a digital link between the telepresence
system controller and the robot controller which enables synchronised time
sampling of the position data to be performed. Thus at any given period in
time the direction of view of the operator and the relative target position can
both be assessed.

The telepresence camera system is able to view an area of approximately
700 X 900mm on the target plane, which covers a total area of 1400 X
1500mm and the target is constructed from a spherical object with a
diameter of 50mm. Due to the relatively wide field of view of the camera
system in comparison with the size of the target, a point of reference
between the vision system and the target being tracked must be established.
The point of reference is required in order to give a reference point from which
measurements can be taken and is formed by overlaying cross-hairs on the
video generated by the vision system. The operator is then required to
maintain the cross hairs on the target for all of its trajectory and the co-
ordinates of the target and the corresponding direction of the line of sight can then be recorded and the characteristic performance levels for different input devices to the telepresence system to be evaluated. To perform this experiment requires the robot, telepresence system and associated equipment to be arranged as shown in the schematic of figure 7.6.

![Figure 7.6 Schematic of Experimental Configuration.]

The telepresence system is arranged such that its look forward orientation is perpendicular to the plane on which the robot is to move the target. The telepresence controller, displays and input devices are located approximately 10m away from the robot and remote telepresence system, such that the operator can not directly view the robot. A description of the configuration of the individual components will subsequently be developed in further detail.
The measurements which can be taken from the telepresence system consist of only the pan and elevation direction of the line of sight. Consequently there is a one to many relationship established along the line of sight vector. That is, for every combination of pan and elevation of the telepresence head there are an infinite number of locations that this can represent in three dimensional space, assuming that the cameras have not been calibrated (Pretlove 1993). Hence, the depth of the target cannot be realised directly from the measurement of the pan and elevation angles only. This can be simply visualised by considering a perpendicular motion of the target, such that the distance between the camera system and the target is increased, along the operators direction of view. Although, the change in depth of the target will be realised by the operator due to the stereoscopic display system, the orientation of the telepresence system will remain constant.

In order to establish the position in space of the target from the orientation of the telepresence system, the target motion is constrained to move on a two-dimensional plane and this plane is arranged to be perpendicular to the forward looking camera direction. To achieve this the robot is moved along a constant depth plane in front of the camera system. When the robot reaches the extreme of its workspace, along this plane, a distance of 2m is measured and marked out from each of the limits. The camera system is then placed at the bisector of these two marks and the direction of the cameras arranged such that they are perpendicular to the marked line.

The target trajectory is measured in terms of the robot co-ordinate system and the line of sight vector is determined by considering the orientation of the telepresence system in terms of its own co-ordinate frame. To evaluate the target tracking ability of the three input devices it is necessary that the
measurements are all with respect to the same co-ordinate frame. This requires calibration of the two co-ordinate frames and the generation of a co-ordinate transform which determines the relationship between the telepresence and the robot co-ordinate frames.

Techniques for recovering the depth of the target plane from the orientation measurements of the telepresence system and the formation of the homogeneous transformation between the two co-ordinate frames are presented in detail.

7.3.2 Position Sampling Synchronisation.

As previously detailed in Chapter 5, the control system of the telepresence system is based on a standard PC processor with additional peripheral interface cards, including a digital input and output card (DI/O). The Unimation Puma 560 Mark III industrial robot is generally configured with a standard VAL II controller. However, for the purpose of these experiments the robot was configured with an additional open architecture controller based on transputers and processors of the 68000 series.

Two robot controllers were used to allow one controller, the VAL II controller, to perform trajectory control of the robot and the other controller to monitor the robot position. Although the VAL II controller is very capable of performing both of these functions simultaneously, it proved very difficult to interface to due to its relatively old architecture. Consequently the second controller was introduced, which offered a simpler method of interface via a MC68030 digital input/output board.
The open architecture controller was designed (Chen 1994) as a layered controller. This allows the closed loop control, trajectory planning and actuation layers to be removed without any detrimental effect on the remaining performance of the controller. The controller can then be used solely for interrogation of the robot encoders and to perform the inverse kinematics to evaluate the Cartesian position of the end effector or in this case the target.

To synchronise the sampling of the target position and the orientation of the telepresence system a digital link was formed between the DI/O board of the telepresence controller and the DI/O board of the open architecture controller. The aim of this digital link was to allow one of the controllers to act as a clock source, initiating a digital pulse at a specific rate. This digital trigger would then force the robot controller and the telepresence controller to record their position and orientation respectively and for these experiments the telepresence controller is used as the pulse source.

The telepresence controller was re-configured to output a digital pulse from the DI/O board under certain conditions. A suitable sampling rate of 0.1 seconds was decided upon to sample the two sets of positions, which is equivalent to fifty cycles of the telepresence controller and over a period of 90 seconds would generate 900 co-ordinate pairs. Consequently, the controller was configured to output a digital pulse to the robot controller every fifty control cycles.

In order to ensure integrity of the digital link a further digital line was introduced between the two control systems. This was required to ensure that the robot controller would only store its position over the same time frame as the telepresence controller. The robot controller was configured to
only record its position when the second line was in a logic 'high' state as shown in figure 7.7.

This line was only set to high by the telepresence controller for the duration of the experiment and then set low at the end to signal the robot controller to cease recording the robot position and notify it of the end of the experiment. The orientations of the telepresence system and the corresponding robot positions were both saved to ASCII files for further off-line analysis.

![Diagram of digital link between telepresence controller and robot controller]

Figure 7.7 The Digital link Between the Telepresence Controller and the Robot Controller.

7.3.3 The Graphical Overlay.

To evaluate the target tracking performance of the system it is necessary to form a reference point between the vision system and the target. A suitable method of achieving this is to overlay a crosshair on to the live video. The crosshair is overlaid on to both video channels of the telepresence system.

The crosshair is generated using a virtual reality (VR) modelling package, known as World Tool Kit, which is running on a separate PC. The VR package is a library of functions which allows three dimensional worlds to be developed and manipulated in real time. The VR system allows the position...
of the cross to be manipulated in three dimensional space relative to the user's viewpoint. The software was configured to generate two channels, left and right, of graphical output and these images were controlled such that the separation of the two crosshairs was manipulated in order for the two crosses to appear as a single three dimensional image when viewed through the stereoscopic display.

Figure 7.8 Left and Right HMD Views.

The cross is generated on a black background to enable colour masking techniques (Kalawsky 1993) to be used to overlay the graphics onto real time video. This is achieved in practice through the use of specialist hardware which mixes the live video and graphics into a single signal which can be displayed. A sample of the images viewed by an operator through the
head mounted display system are shown in figure 7.8. This shows the crosshairs and the target mounted on the end effector of the robot.

Once the cross is set in position and a clear stereoscopic image formed, it remains set in this position for the duration of each set of experiments. The display system is configured as shown in figure 7.9. This figure shows the configuration of the display systems for both immersive and non-immersive viewing of the graphical overlay and video.

![Display System Configuration Diagram]

Figure 7.9 Display System Configuration.

7.3.4 Development of the Governing Equations.

The position of the target and the orientation of the telepresence system can now be sampled in synchronisation, through the use of the DI/O link.
However these measurements are made in different frames of reference, the robot position is measured in the robot world frame and the telepresence orientations are measured in the telepresence co-ordinate frame. Therefore in order to evaluate the tracking performance it is necessary to establish the relationship or transformation between the two co-ordinate frames. This requires the calibration of the telepresence frame to the robot co-ordinate frame.

The aim of the governing equations is to establish the position of a given point in one of the reference frames in terms of the other reference frame. The transformation between the telepresence frame of reference and the robot co-ordinate frame will therefore be established. The equations are determined using vector equations of lines and homogeneous co-ordinate transformations.

Figure 7.10 Target and Telepresence System Configuration.
Figure 7.10 shows a schematic of the configuration of the telepresence system and the robot. The aim is, from a given set of telepresence system orientations, to calculate the position where the operator is looking in terms of the Cartesian location in the robot co-ordinate frame.

### 7.3.4.1 Determining the Depth of the Target Plane.

Consider a vector from the origin of the robot co-ordinate frame (RCF) and the line of sight vector from the origin of the telepresence co-ordinate frame (TCF) to the target, as shown in figure 7.10. The vector to the target from the origin of the RCF defines a location in the robot co-ordinate frame in terms of the Cartesian position given by the robot X, Y and Z locations. Now consider a vector from the TCF origin to the target, this represents the direction of view of the operator.

The measurements which can be taken from the telepresence system do not define a specific location in space, they only describe the orientation of the line of sight vector in the horizontal and vertical planes. Hence, the distance along the line of sight vector cannot be recovered by taking measurements of the orientations alone. By constraining the target motion to lie in a single plane, a mathematical solution can be formed in order to determine the location of the target along the line of sight vector.

The motion of the target is to be constrained to a single vertical two dimensional plane in the robot co-ordinate frame, the effect of which is twofold; (i) it prevents any changes in depth of the target occurring which gives rise to a zero change in the pan and elevation orientations of the telepresence system and (ii) it aids the recovery of the viewpoint location along the line of sight vector by constraining it to be located on this plane.
As shown in figure 7.11 the plane of the target motion is at some unknown distance, $Y_t$, from the origin of the telepresence co-ordinate frame. In order to resolve the orientation measurement and calculate the position of the viewpoint with respect to the TCF it is necessary to determine the distance, $Y_t$. However this requires the TCF and the RCF to be arranged such that a constant depth vertical plane in the TCF is parallel to a constant depth vertical plane in the RCF. This can be achieved in the experimental
configuration by simply ensuring that the telepresence system, whilst in its 'look forward' position, is perpendicular to the plane of motion of the target. This is arranged as part of the experimental configuration as detailed in section 7.3.1

The perpendicular separation between the target motion plane and the TCF origin can be determined by considering the equations of the vectors to two points of a known separation in the target plane, \( \Delta \), shown as \( P_1 \) and \( P_2 \) in figure 7.11. The vector equations of these two lines are of the form given by equations 7.1 and 7.2.

\[
\begin{align*}
P_1 &= \lambda_1 \begin{pmatrix}
\cos\phi_1 \sin\theta_1 \\
\cos\phi_1 \cos\theta_1 \\
\sin\phi_1
\end{pmatrix} \\
P_2 &= \lambda_2 \begin{pmatrix}
\cos\phi_2 \sin\theta_2 \\
\cos\phi_2 \cos\theta_2 \\
\sin\phi_2
\end{pmatrix}
\end{align*}
\]

The equations, 7.1 and 7.2, describe the vector to a point \( P_1 \) for a given \( \theta_1 \), \( \phi_1 \), and elevation angles of the telepresence system. If the two points, \( P_1 \) and \( P_2 \), are of a known separation \( \Delta \) and are both located on the same plane then the following conditions are established.

A vector, \( d \), joins the points \( P_1 \) and \( P_2 \) which yields

\[
d = P_2 - P_1
\]
or alternatively,

\[
d = \begin{pmatrix} P_{2x} \\ P_{2y} \\ P_{2z} \end{pmatrix} - \begin{pmatrix} P_{1x} \\ P_{1y} \\ P_{1z} \end{pmatrix}
\]

but the two points are located in the same vertical plane so,

\[P_{2y} - P_{1y} = 0\]

A further constraint can be introduced to simplify the vector manipulation by assuming that the points \(P_1\) and \(P_2\) are located on the same horizontal plane with a known separation, yielding,

\[P_{2x} - P_{1x} = 0\]

and

\[P_{2z} - P_{1z} = \Delta\]

In practice this proved very easy to attain as it simply requires driving the robot to two points where the robot \(Y\) and \(Z\) co-ordinates are maintained at constant values.

Consequently the vector \(d\) can be re-written as

\[
d = \begin{pmatrix} \Delta \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \Delta \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}
\]

These conditions allow equations 7.1. and 7.2 to be manipulated, as shown in Appendix E, to yield an equation for \(\lambda_1\) in terms of the orientation angles, \(\theta_i\).
Chapter 7: Task Performance Evaluation.

pan, and \( \phi \), elevation, the separation of the two points, \( P_1 \) and \( P_2 \), \( \Delta \) as shown in equation 7.3.

\[
\frac{1}{\lambda_1} = \frac{1}{\Delta} \left[ \frac{\sin \phi_1 \sin \phi_2}{\tan \phi_2} - \cos \phi_1 \sin \phi_1 \right] \tag{7.3}
\]

Substitution of equation 7.3 into equation 7.1 enables the depth of the plane in which the points, \( P_1 \) and \( P_2 \), are located to be determined, yielding equation 7.4.

\[
P_{1y} = \Delta \frac{\tan \phi_2 \cos \phi_1}{\tan \phi_1 \sin \phi_2 - \tan \phi_1 \tan \phi_2 \cos \phi_1} \tag{7.4}
\]

So, for any given pair of points which satisfy the above conditions, the horizontal distance between the TCF origin and the target motion plane can be resolved from a measurement of the pan and elevation angles of the telepresence system. Then by using simple trigonometry, as shown in Appendix E, the location of the viewpoint can be specified in terms of the telepresence system Cartesian \((T_x, T_y, T_z)\) co-ordinate frame.

The target motion is to be constrained to a trajectory in the target plane, consequently the depth evaluated from equations 7.3 and 7.1 can be used for resolving all the viewpoint locations throughout the experiment which in effect constrains the viewpoint to be on this plane.

7.3.4.2 Determination of the Transformation Matrix.

The distance between the telepresence co-ordinate frame (TCF) and the target plane is established as the initial stages of a calibration routine prior to the experiments. This however only determines the location of the
viewpoint in terms of the TCF. In order to perform a direct comparison of
the tracking performance, the viewpoint location must be transformed to
reference the same co-ordinate frame as the target trajectory. This can be
achieved through applying a homogeneous transformation to the TCF.

A homogeneous transformation matrix is a 4x4 partitioned matrix with the
partitions representing rotations, translations, scaling and perspective
transformations. A detailed examination of which can be found in many
robotics texts (Stadler 1995). The transformation is applied to the viewpoint
locations to yield the corresponding Cartesian position with respect to the
robot co-ordinate frame, as shown in equation 7.5.

\[
\begin{bmatrix}
C_{11} & C_{12} & C_{13} & C_{14} \\
C_{21} & C_{22} & C_{23} & C_{24} \\
C_{31} & C_{32} & C_{33} & C_{34} \\
C_{41} & C_{42} & C_{43} & C_{44}
\end{bmatrix}
\begin{bmatrix}
T_x \\
T_y \\
T_z \\
1
\end{bmatrix}
= \begin{bmatrix}
R_x \\
R_y \\
R_z \\
1
\end{bmatrix}
\] (7.5)

where \(C_{ij}\) are the characteristic parameters of the transformation,
\((R_{xi}, R_{yi}, R_{zi})\) and \((T_{xi}, T_{yi}, T_{zi})\) are the target and viewpoint co-ordinates
respectively.

In this case the perspective, \(C_{41}\) to \(C_{43}\) and scaling factor, \(C_{44}\), components
of the transformation can be set to zero and unity respectively. Thus
reducing the number of unknowns. As shown in Appendix E, equation 7.5 can
be re-arranged to give three equations in terms of the twelve unknown
transformation coefficients, the target locations and the viewpoint locations.
By considering a set of twelve target and corresponding viewpoint positions
the three equations are over determined and a least squares solution for the
\(C_{ij}\) coefficients can be performed. In terms of matrices this can be written as
shown in equation 7.6.
The equation 7.6 can then be solved for the transformation coefficients using Matlab (Herskovitz 1989) to calculate the pseudo inverse of matrix \([T]\). The components of the resultant \([C]\) matrix are then known as the calibration coefficients. These calibration coefficients then form the homogeneous transformation which is used to transform the Cartesian viewpoint locations from the TCF to the robot co-ordinate frame (RCF).

In using a least squares fitting approach to solve for the calibration coefficients it is necessary for the \([T]\) matrix to exhibit full column rank. To ensure this occurs, it is necessary to utilise a further six points which are located in a different plane. This is performed by performing a second calibration routine using a set of points located in a plane separated by a known distance from the first plane.

Therefore in order to compare the target tracking performance it is necessary to perform a two stage calibration. The first stage is to determine the depth of the target plane from the TCF origin using two points in the plane with a known separation. The second stage is to determine the homogeneous transformation between the TCF and the RCF using an array of six points in the target plane and six points in another parallel plane.
Ideally each of the groups of six points should be distributed over the plane such that the calibration is interpolated over the plane. Once the homogeneous transformation has been determined the Cartesian position of the viewpoints and the target trajectory can be compared with respect to a common co-ordinate frame, which is in this case the robot co-ordinate frame.

### 7.3.5 Experimental Procedure.

To assess the target tracking performance of the various input devices requires the target to be moved at a range of speeds which is governed by the robot speed, and is easily configured in software on the robot controller. The tracking experiments are performed for each input device over a range of robot speeds. The robot speed is set as a percentage of the maximum joint speed and the experiments are performed for maximum robot joint speeds of 10, 20, 30, 40 and 50 percent.

The calibration routines discussed in section 7.3.4 are performed prior to each experiment to ensure the correct calibration data is used. The telepresence system orientations and the robot position are also stored in separate files on the telepresence and robot controllers respectively for each experiment.

Each experiment was performed by four different people who have no experience in the use of telepresence systems and each person was allowed a training period of two minutes prior to the commencement of the experiment, during which the target was maintained in a stationary position. The operators were then requested to track the moving target for 90 seconds, maintaining the cross hairs on the centre of the target for each of the input
devices. All of the experiments utilise a head mounted display as the display device.

For each operator the experiments were performed in the same order, such that any learning effects should be common to all of the tests.

7.3.6 Experimental Results.

The initial stages of the target trajectory tracking experiments involve determining the homogeneous transformation between the telepresence head co-ordinate frame and the robot co-ordinate frame. This transformation allows the Cartesian co-ordinates of the point at which the operator is looking on the target plane to be resolved from only a measurement of the orientation of the telepresence camera system. Subsequently this viewpoint location, in terms of the telepresence co-ordinate frame, can be expressed in terms of the robot co-ordinate frame. Figure 7.12 shows the position of the robot locations used in the calibration process and the corresponding viewpoint locations established during the calibration process, transformed into the robot co-ordinate frame. These points have been calculated through applying the homogeneous transformation evaluated using the techniques discussed in section 7.3.4. The view points shown in figure 7.12 are the average viewpoint positions determined during the calibration of the head coupled, Polhemus joystick and digital joystick controlled systems. The RMS error associated with each calibration point is also shown in figure 7.12 and approximates to an overall average RMS error of 75mm. The RMS error is evaluated by using equation 7.7 and the average RMS error for all the calibrations is calculated using equation 7.8.
where $V_{xi}$, $V_{zi}$ are the X and Z co-ordinates of the viewpoint and $R_{xi}$, $R_{zi}$ are the target co-ordinates at the corresponding sample time. This signifies that all of the calculated co-ordinate locations will be in error by an average of 75mm. However, all of the calculated target trajectories will be subjected to this error and consequently for a comparative study the effect is less significant. It is the relative differences between the trajectories which is of interest and not the absolute values of the trajectories.

Figure 7.12 Mean Calibration Error.
The calibration routines are performed for each operator and input device and the corresponding transformation matrices evaluated in each case. The operators then track the target trajectory generated by the robot.

Figures 7.13 to 7.16, show the target trajectory and the corresponding trajectory of the operator viewpoint for one typical operator. The viewpoint trajectory has been calculated by applying the homogeneous transformation to the location of the viewpoint, in terms of the telepresence co-ordinate frame. The traces shown relate to target speeds of 10 and 50 percent of the maximum robot joint speed for each of the head coupled and Polhemus joystick controlled test systems respectively. For each of the traces the time period under consideration is not the same for the 10 and 50 percent robot speeds, but for each of the given robot speeds the time periods are identical. This is because the target moves a greater distance in the same time period as the robot speed is increased.

![Figure 7.13 Polhemus Joystick Tracking at Target Speed of 10.](image)

Figure 7.13 Polhemus Joystick Tracking at Target Speed of 10.
Figure 7.14 Polhemus Joystick Tracking at Target Speed of 50.

Figure 7.15 Head Coupled Tracking at Target Speed of 10.
To enable a comparison to be made between the four operators and the various input devices used in the experiment, the RMS error has been calculated for each of the tracking tasks, using equation 7.7. The task was to track the target using three types of input device and a range of 5 robot joint speeds between 10 and 50 percent. Figure 7.17 shows the RMS error for each operator and input device for the minimum and maximum target speeds. This summarises the RMS errors generated by each operator and allows general trends to be clearly seen within each operator group.
Figure 7.17 Summary of High and Low Speed Target Tracking.

For each of the operators the digital joystick input device exhibits the largest error at the highest robot speed. At the lower robot speed the head controlled and the Polhemus based joystick input systems offer a comparable tracking error. However as the target speed is increased, so too is the RMS error for the Polhemus joystick input device. The results of testing with each individual input device are given in figures 7.18 to 7.21 for the head, Polhemus joystick and digital joystick controlled systems respectively.

These figures clarify the information contained in figure 7.17, showing that the tracking errors tend to increase within each type of input device, as the target speed is increased. It also shows that there is a greater significant increase in the RMS error for the Polhemus based joystick system as the robot speed is increased, than for the head controlled system.
To determine the overall performance for this experiment and to allow a further comparison of the three input systems to be made, the average RMS error over the range of robot speeds has been evaluated for each of the devices. This is shown as three curves, one for each of the devices, in figure 7.22.

Figure 7.22 shows that for a target tracking task over a range of target speeds the head coupled telepresence system offers better tracking.
accuracy. For three out of the four operators, the head controlled system performed better than the Polhemus based joystick and the digital joystick. In all cases the digital joystick exhibited very large tracking errors.

![Average RMS Target Tracking Error](image)

Figure 7.22 Average RMS Target Tracking Error.

### 7.3.7 Discussion of the Experimental Results.

The target tracking experiments show that, for this particular type of task, the use of a head coupled system is able to offer an improved operator performance. This is particularly noticeable at high target speeds where the use of a simple digital joystick is prohibitive due to its appreciable reduction in operator performance.

The trajectory following capability of the head coupled and Polhemus joystick input devices are shown in figures 7.13 to 7.16. These traces show how a particular device follows the trajectory of the target. It can be seen that for the higher robot speeds there is a significant amount of overshoot of the viewpoint trajectory during a change in direction of the target. This effect is
far more apparent with the Polhemus and digital joystick based input systems.

At the higher robot speed, although there was a notable degree of overshoot with the joystick input systems, it did appear that the operator had a tendency to simultaneously move their head in the direction of the target whilst using the joystick input. In fact this was also a very common occurrence with many of the middle to higher range of target speeds. This could be accounted for by the operator experiencing a sensation of their vision being directly linked to the position of the cross hairs due to the use of the head mounted display unit. Consequently this seemed a natural motion to the operator and must reflect a degree of visual immersion in the remote scene.

During tracking the target at the slower robot speeds the head controlled system exhibits a much smoother trajectory throughout the complete target path. From the plots of the target trajectory it is possible to notice a degree of searching along the Polhemus joystick trajectory which could be explained by the sensitivity of the joystick. Although this was ensured to be identical to the head controlled system, it may have not been an optimal configuration for a joystick. The digital joystick simply exhibits a much coarser trajectory path and exhibits difficulty in maintaining a smooth trajectory at any speed which is lower than the configured 'active' speed of the joystick.

The trajectories shown in figures 7.13 to 7.16 do not allow any context of time to be deduced. It may be possible for the viewpoint trajectory to be lagging behind the target but still appear from these figures to be completely on target. This effect is quantified by the calculation of the RMS error.
between the target and the viewpoint paths, shown in figures 7.17, 7.18, 7.19, 7.20 and 7.21.

For each individual operator the RMS error exhibited during low target speeds, when using the Polhemus joystick and the head controlled system, were very comparable. However, as the target speed increased the head controlled system maintained a lower tracking error.

Although the tracking errors are comparable at the lower speeds the operator still moved their head in the direction of the target motion, even when using one of the joystick input devices. Consequently it will be argued that the head controlled system offers a more natural interface for performing this type of task, especially as the speed of the target motion is increased. It is probably true, however, that for a manipulation type of task that a joystick would be better suited for controlling the manipulator and the head controlled system for controlling the viewing system as this is a more natural interface arrangement. This is a configuration similar in concept to the experiments performed in section 7.2.

The target trajectory used in all of the tracking experiments was designed such that it was very difficult to predict the future position of the target over a period of time. However, as the trajectory was not purely random, which would have allowed virtually zero prediction to be employed, the effect of learning on the results of the experiments cannot be ignored. To prevent any biasing of the experiments due to operators learning the trajectory, all of the experiments were performed in exactly the same order with each operator. This ensured that if the operators were able to learn the target trajectory that the effect was replicated throughout all of the experiments. It was, in fact, not until the later stages of the experiments that any of the operators
realised that an identical trajectory had been used for each test. This is probably due to constantly changing the target speed and the input device.

During certain sections of the target trajectory it was possible for the target to remain in the operators field of view without the necessity of any head motion. This was anticipated prior to the experiments and was accounted for by using the cross hairs as the tracking reference point and not the operator's field of view. Consequently, any motion of the target required motion of the operator's head in order to maintain the cross hairs on the target.

The evaluation of the target tracking performance of the three input devices relies upon the calibration of the telepresence co-ordinate frame to that of the target or robot. This requires the operator to accurately locate the cross hairs of the display on the centre of the target. The corresponding orientations of the telepresence head must then be measured. This proved especially difficult in the case of the Polhemus based joystick, as it was very difficult to keep the camera system perfectly stationary whilst the necessary measurements were taken. It also relies on the operator accurately locating the cross hairs onto the centre of the target.

Any errors which occur during the calibration stage will be propagated through to the final results which could explain the variance in the RMS errors which are calculated for the same test with different operators. It is also the case that there should be some natural variance due to some operators performing the tasks much better than others, especially if they exhibit superior hand to eye co-ordination.
There is also a further error introduced into the system which is in the form of a modelling error. This is introduced when modelling the telepresence head as a three dimensional co-ordinate system with a common origin for rotations in the horizontal and vertical planes. The actual system is designed with the horizontal axis of rotation offset from the vertical rotation axis by 50mm. An estimation of the errors introduced by this approximation are developed in Appendix F.

Over the range of motions experienced during the experiments, a maximum error of 58mm is introduced at the extremes of the target trajectory. However, it is a comparison of the RMS errors which is of interest and not the absolute error which is of importance. All of the results can be assumed to be subjected to the same magnitude of this modelling error. Consequently, the affect of this error on the comparative results of these experiments is negligible.

7.4 An Assessment of the Effects of Latency.

The latency of the telepresence system has been evaluated, as shown in section 6.3, to be 78ms when using a Polhemus system as the demand input to the system. These experiments aim to show the effect of increasing the latency period on the performance of the telepresence system for target tracking. Investigations of the effects of the magnitude of the system latency period were performed in order to demonstrate the necessity for a low latency period which is required to satisfy operator comfort and task performance requirements.

The experiments use the same calibration theories as that of experiments performed in section 7.3, but only use the head tracked form of the Polhemus
system as the input device. All of the tests use the HMD display as the viewing device and the experimental configuration is also as shown in figure 7.6.

These experiments do not aim to establish the physiological aspects of the effects of latency on operator performance. It is to quantify the effects on target tracking performance experienced through a change in the latency period of the telepresence system. The physiological effects are no less important, but their detailed investigation is beyond the scope of this experiment. However, some general comments made by the operators during and after the experiments will be presented and discussed.

The experiments required the operator to track a moving target, which was again moving with a pseudo random trajectory. The experimental and display system configurations were as shown in figures 7.6 and 7.9 respectively.

Unlike the previous experiment where the robot speed was varied, this experiment is performed with a constant mid-range robot speed of 30 percent of the maximum joint speed. The experiment is performed using this robot speed for a range of latency periods which range from 78ms, the system minimum, to 2000ms.

The system was designed with a minimal latency period in mind, however due to the open architecture design of the control system, the latency period can be increased from this minimum value very easily in software. The increase is performed by essentially buffering the input from the input device for a given period of time. This takes place on the telepresence controller and can be configured to generate effectively any latency period greater than the system minimum.
The experiment was performed by four people who had no experience in using the telepresence system. They were given a two minute training period, where the target was stationary, in order to become accustomed to the system and its behaviour. For the training period, the minimum system latency was used.

As with the experiments in section 7.3, the calibration routine was performed prior to each set of experiments to evaluate the calibration coefficients. The operators were then asked to track the target for a period of 90 seconds. The robot and telepresence system positions were recorded every 100 ms and stored to file for further analysis.

### 7.4.1 Experimental Results

Once the calibration has been performed the homogeneous transformation matrix can be established and the trajectory of the viewpoint determined. Following the evaluation of the viewpoint trajectory in terms of the robot coordinate frame the RMS error for each of the latency periods can be determined. The RMS tracking error value for each of the latency periods considered in the experiment are shown in figure 7.23.

Figure 7.23 clearly shows that the RMS error increases as the latency time increases. This signifies that the operators experienced greater difficulty tracking the target with a large latency period.
7.4.2 Discussion of Latency Experimental Results.

From figure 7.23, it can be stated that the minimum RMS error for the target tracking task occurs with the minimum latency period of 78ms. There is a sharp initial increase in the RMS error as the latency period increases beyond 100ms and then a steady increase in error as the latency rises towards 1000ms.

During the experiments, a number of observations were made regarding the techniques adopted by the operators to track the target trajectory. At the minimum latency period, the operators were observed to perform a very smooth trajectory following motion. However, at the larger latency periods the operators often lost complete sense of the orientation of the camera system. In some cases the operators were observed to remain physically
stationary until the telepresence system caught up, in order to try and regain control and establish their direction of view.

At high latency periods the viewpoint trajectories were also very oscillatory, with a significant degree of overshoot, as the operators attempted to gauge how far they needed to move their head in order to stay on the target path. This was not an attribute observed during target tracking whilst using the minimum latency period.

During the experiments many operators complained of experiencing a high degree of disorientation with latencies in excess of 1400ms. Although a similar sensation, but to a lesser extent, was also perceived at latencies over 400ms. When the latency was increased to 2000ms, all of the operators felt sure that they would soon feel the symptoms of motion sickness if subjected to the system for a prolonged period of time.

As the latency periods increased in excess of 400ms, it became more and more difficult for the operators to accurately control the camera system. Further increases of the latency simply enhanced the difficulty of controlling the system. When the latency period was less than 400ms the operators would feel that the system would respond to their demands, but with latencies in excess of this level a sensation of physical separation from the camera system would be experienced. At this level of latency the sensation of telepresence is significantly reduced.
Chapter 7: Task Performance Evaluation.

7.5 An Experimental Comparison of Immersive and Non-Immersive Display Systems.

These experiments aim to compare the operator performance during a target tracking task when head mounted immersive and monitor based non-immersive display systems are used. It involves performing the same task as described in section 7.3 but only using the Polhemus based joystick as the input device. Through using a common input device the relative merits of the HMD based system and the monitor based display system can be investigated for this particular application.

To perform these experiments a Tektronix stereoscopic display device is used. This display unit uses the principle of temporal sequencing, a technique which is discussed in detail in section 3.5.2.2. The experimental configuration is shown in figure 7.6, with the Tektronix display and the associated driving hardware introduced into the overlaid video circuit, as shown in figure 7.9. When the operator is observing the stereoscopic display they must wear a pair of polarising glasses and for the optimum stereoscopic effect are requested to view the monitor from a distance of 0.5m.

The experiment consists of the operator performing a series of target trajectory tracking tasks, using the Polhemus based joystick as the input to the telepresence system. The target pseudo random trajectory was again generated by the Puma Industrial robot and the relevant sampled target positions and telepresence orientations stored to file. Only a single mid-range robot joint speed, of 30 percent, was used for these experiments. Prior to the experiment each operator was allowed a training session of two minutes when a static target was made available to view and the latency period set...
to the minimum level. Four operators were used for these experiments and two sets of experiments were performed with each operator. The first set utilised the head mounted display unit and the second utilised the monitor based stereoscopic display. For each of the display systems, the operators were asked to perform the calibration routine and then track the target as closely as possible for 90 seconds.

7.5.1 Experimental Results.

Again, once the calibration matrix was determined for each series of tests, the viewpoint trajectory could be calculated and the RMS tracking error established. The RMS tracking errors of the experiments, using both the non-immersive and immersive display technologies, are shown in figure 7.24. The aim of these experiments was to perform a comparison of non-immersive and immersive display systems in their simplest form. They show, for the target tracking task, that the operator performance is slightly improved through the use of an immersive display system.

![Figure 7.24 A Comparison of the Target Tracking Error Using Immersive and Non-Immersive Display Systems.](image_url)
7.6 Conclusions.

This chapter aimed to evaluate the application of the telepresence system to a number of operator oriented tasks. This was achieved by a series of experiments which included; an assessment of the advantages offered by stereoscopic viewing and an investigation of the target tracking capability of the telepresence system. The operator oriented tasks required operators to perform a range of experiments with some distinct differences in the experimental configuration. The environment used to perform the stereo advantage experiments required the operators to use the telepresence system as a visual interface to a telerobotic cell, which demanded interaction between the operator, the telepresence system and a robot manipulator. The second area of experimentation required the operator to only control the telepresence system in order to interact with a rapidly moving target.

The experiments used to demonstrate stereo advantage showed that a stereoscopic viewing system enhanced the operators spatial perception of a remote environment. This was demonstrated by a reduction in the task completion time, over the monoscopically observed case, and improved operator confidence during the determination of the location of the telemanipulator in three dimensional space. During these experiments the operators were observed to constantly search the workspace in order to plan the future motion through which the manipulator should move. This seemed to occur naturally due to coupling of the camera system motion to the operators head motion.

These experiments demonstrate that where direct viewing of the workspace is not possible, stereoscopic viewing allows three dimensional tasks to be performed more quickly and accurately when compared to a monoscopic
viewing system. The limitation of these experiments was that they didn’t compare the performance whilst using a joystick interface to control the telepresence system, but from the observed behaviour of the operators the head coupled system presented an ideal interface in order to allow the operator to visually interact with the remote scene. This allowed the operator to focus on the task of operating the robot manipulator and to be less concerned with controlling the remote camera system. These experiments determined that the head coupled form of the telepresence system is an ideal interface in this instance, and when combined with stereoscopic viewing presents an improved technique for performing remote telemanipulation tasks.

The second area of experimentation focused on the assessment of target tracking capability and involved many variations of the same experiment with different experimental constraints. Essentially these experiments evaluated the target tracking error of the operator controlled telepresence system. Three types of the tracking experiment were performed to investigate (i) the type of input device to the system, (ii) the effect of latency on tracking ability and (iii) an investigation of how the format of the display device affects tracking performance.

In order to perform all of these three experiments a technique for recovering the location on a target plane where the operator is looking via the telepresence system, known as the viewpoint, from a measurement of the orientation of the camera system was developed. The determination of the viewpoint from the orientations of the telepresence system required a number of assumptions and constraints to be made regarding the experimental configuration. In most cases the constraints, such as ensuring that two of the calibration points lie on the same horizontal and vertical
planes, are quite easily and accurately achievable, the limitation to this being the requirement for the look forward position of the telepresence system to be perpendicular to the target plane. This required manual measurement and alignment of the telepresence system which is prone to human error. Using these constraints and assumptions the accuracy of the viewpoint estimation demonstrates a RMS error of 75 mm over the workspace. The main contribution to this error is due to the simplistic model adopted for the telepresence system which introduces a mean RMS error of 50 mm across the workspace.

Although this error affects all the experiments, each one is a comparative study and consequently all the results are affected by the same amount and so should not demonstrate any effects to the underlying trends which have been shown.

Experiments were performed to investigate the target tracking ability whilst using a number of different operator input devices. The experiment investigated three forms of operator input via a digital joystick, a Polhemus based joystick and a Polhemus based head tracking system. The head coupled system demonstrated a reduced RMS tracking error when compared to the joystick systems and this was especially evident with a quickly moving target. The Head coupled system has shown to offer advantages when the target was moving quickly or when rapid motion of the operator's head was required. This is due to the head coupled system presenting an improved and natural interface to perform control of the remote camera system.

The second range of experiments, using the same target tracking methodology, investigated the effects of the magnitude of the latency period
on the target tracking ability of the operator. These experiments aimed to focus on the quantitative effects of the latency period on the tracking accuracy with latency periods in the range of 80 - 2000 ms. In situations where the latency period was in excess of 400 ms, there was a significant degradation in the performance of the target tracking task. This was due to the operator experiencing sensor conflict, signals presented to the operators vision system opposed those generated from the operators head motion and consequently introduced a lack of correspondence between the operators head motion and the remote camera system. It is the time delay between the operators action and the re-action of the remote system which causes these conflicts to occur. This is one of the causes of motion sickness and is also very detrimental to the sensation of remote visual presence. Consequently these experiments have demonstrated that for a high degree of operator immersion and comfort during tasks which require rapid interaction, a low latency period must be utilised.

The telepresence system used in the experiments has a low latency period of around 80ms which allows the operator to sense a seamless link to the remote camera system. However, attempts to further reduce the latency period could be well justified, especially when the system is used in an environment where latency effects due to other components are unavoidable, such as during the use of radio links for a wire-less link between the operator and the remote environment.

Although the physiological affects of increased latency periods were not intended to be the focus of the experiments a number of issues were observed and raised by the operators. At higher latencies, in excess of 400 ms, the remote system sometimes appeared to be out of control and the operator
would lose all sense of the camera orientation and afterwards commented on the expectation of prolong use to induce discomfort.

The final set of experiments compared the use of a non-immersive and immersive display systems, with the telepresence system, to perform the same target tracking task. The experiments showed that the head mounted display system allowed the operator to perform the tracking task slightly better than when using the non-immersive display. This can be attributed to the operator being highly visually immersed in the remote environment. It is expected that as HMD technologies further improve, with respect to resolution and field of view, that the distinction between the performance of the HMD and non-immersive display will be further increased. This experiment used the same joystick input device to control the camera system, so that the performance using the different display technologies could be contrasted. It is expected that if the HMD was used with the head-coupled form of the system that the performance would be further improved as it is not only the visual sense that is immersed but also the motor actions of the operator's head, thus increasing the sensation of remote presence.

7.7 Summary

The experiments performed in this chapter have validated the design and performance of the telepresence system. It has been shown to enhance operator spatial perception of a remote environment and to improve the performance of a target tracking task by demonstrating a minimal latency period.
CHAPTER 8

CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

8.1 Introduction.

Telepresence systems are an essential aspect of teleoperation and remote inspection systems. Their development cannot be performed without a significant consideration of the human factors which dictate the performance requirements of the system. Until such a time when autonomous robotics are capable of performing complex tasks extremely well in unstructured, unknown and often dangerous environments, telepresence systems will still be required.

They offer the possibility of removing an operator from a potentially hazardous situation, whilst still maintaining a sense of visual presence at the remote location. The advantage of a full colour three dimensional representation, when presented to an operator in such a way that they can visually interact with the scene as normal, offers numerous advantages over current static systems. The precept of depth will allow far more complex manipulative tasks to be performed with confidence and head tracking provides a natural interface for the manipulation of the visual presentation.

This chapter presents conclusions regarding the research work and also proposes further research which will improve the performance and use of the
telepresence system. It will also suggest alternative applications in which the telepresence system can be utilised.

8.2 Conclusions.

This research work describes the development and evaluation of a remote telepresence system. The system comprises of a four, independently controlled, degree of freedom mechanical head with two ‘state of the art’ colour CCD cameras. The images are displayed to the operator in one of two ways, using either an immersive head mounted display (HMD) or a non-immersive stereoscopic monitor based system.

An open architecture control system has been developed which enables very simple interfacing of external sensors to the system. The system currently utilises a six degree of freedom electromagnetic sensor to provide real-time tracking of the operator’s head motion.

The system demonstrates excellent trajectory following, with a latency period around 78ms. This latency time is almost 50% of the time quoted as the performance of the NASA system (Bolas and Fisher 1990), which can be attributed to the dedicated controller software, interrupt driven communications and the integrated mechatronic design. The latency of a telepresence system is of paramount importance when assessing its performance, as demonstrated by using the system to perform a target tracking task with a range of latency periods between 78ms and 2000ms. At latency periods in excess of 400ms the operators exhibited a sense of disorientation and felt that after prolong use that they would subsequently exhibit the attributes of motion sickness. Thus for such a system to be acceptable for use, the latency effects must be minimised.
Other performance considerations of the system include the dynamic performance when subjected to different types of input request, for example a 90° pan motion can be achieved in under 0.3 sec. The ability of the system to track the dynamic head motion trajectories demanded by an operator were also investigated. The system was shown to be capable of tracking the most rigorous motion profiles demanded by the operator extremely well, whilst exhibiting a very controlled motion.

During experimental application to three dimensional locating and tracking tasks the system has been shown to exhibit significant advantages over monoscopic systems. This is achieved through increased operator spatial perception enabling a reduction of completion time and improved accuracy for a three dimensional locating task. Further, with the ability of the operator to directly and intuitively manipulate the gaze direction of the cameras the available observation space is also increased.

A number of operator performance oriented experiments were performed to evaluate the most suitable form of trajectory demand input to the system. These experiments involved tracking a highly dynamic target whilst using a range of input devices such as joysticks and a head motion tracking system. The head controlled system was demonstrated to exhibit superior tracking capability of high speed targets. The system also enabled an initial investigation into the effects of the system latency period on the operator performance to be investigated. These experiments showed that a low latency period is required in order to attain operator comfort and to provide a high performance telepresence system.

Overall the system, which has been developed and demonstrated, has shown to be of very high performance and can form the basis of further research.
into the benefits of remote telepresence and its effects. Due to the open nature of the controller the system can also act as a test bed for other tracking sensors or sensor development.

**8.3 Future Research Directions.**

This research work has produced a working experimental telepresence system. However there are still some areas which can further enhance the performance of the system and also lead to other areas of indirect application of this technology. Further possible research areas would include the following;

1. **System Modelling.**
   The system model which was developed as part of the design phase does not truly represent the performance of the system as there are significant flaws in the model. These are mainly due to the many non-linear components, such as the power amplifiers and time sampled characteristics of the system. The theoretical system model would be further developed by the introduction of a more complex model, with accurately evaluated parameters, such as the load inertia. Since the system has now been built an alternative would be to use system identification techniques to evaluate a system transfer function. A knowledge of the system transfer function will be useful in order to investigate a more advanced and higher performance control algorithm.

2. **Head Motion - Trajectory Prediction.**
   In comparison with other telepresence systems the latency period is very short. However with the introduction of more advanced control schemes into the control algorithm it is anticipated that this delay could be further reduced. One technique would be to introduce a form of prediction, utilising
least squares or Kalman filter approaches. The other technique would be to develop more complex control routines such as adaptive or feed forward techniques. The necessity for this has been demonstrated by the experiments present in this thesis. Large latency periods essentially doom the system almost unusable and has a significant effect on the comfort of the operator during highly dynamic tasks, such as target tracking. To use predictive control in the control algorithm may enable the system to be used where the introduction of latency effects are unavoidable, such as in situations where a radio link is being used between the operator and the telepresence system.

(3) Latency Investigation.
The experiments assessing the effects of an increased latency period on the operator performance during a target tracking task could be carried out with a focus on evaluating the physiological effects. This is important as it may be possible for the operator to perform the target tracking task adequately well with a large latency period over a short time frame, but after prolong use the physiological effects may be significant.

(4) Mechanical Design.
The introduction of a further degree of freedom about the head roll axis could increase the operator sensation of presence. However, the effectiveness of this addition would have to be evaluated in terms of the increased cost and design complexity. The absence of this degree of freedom from the design of the current system has not had any significant effects or produced any limitations on the work presented in this thesis.
(5) Head Motion Calibration.
During elevation motions of the operator's head the angle which is realised by
the head mounted sensor varies depending on the location of the sensor on
the operator's head. This effect can be minimised by performing calibration of
the operator's head motion to the motion of the mechanical structure. This
calibration will enable trajectory generated by the mechanical system to
account for the differences in the operator head and telepresence system
kinematics.

(6) Intelligent Mobile Telepresence.
An application in which the telepresence system could be utilised is in the
area of intelligent mobile telepresence. A mobile telepresence system
consists of the telepresence system mounted on a mobile vehicle. During the
use of such a mobile telepresence system the operator is subjected to many
different stimuli which either require action or decision. It would be preferable
for the operator to only have to deal with the most important of tasks which
could not be performed autonomously. One method of achieving this is
through human supervisory or shared control. Supervisory control can be
achieved by giving the mobile vehicle a range of on board intelligence such as
a collision avoidance system. Then, through overlaying a graphical virtual
environment on the real scene, it would be possible to control the robot using
virtual reality (VR).

The VR could be used to form the path planning of the vehicle, once selected
the operator would switch to autonomous mode and the mobile would
automatically move to the requested position whilst avoiding any obstacles,
then hand control back to the operator. This will remove the necessity for the
operator to continuously control the vehicle, leaving them free to perform
other tasks via the telepresence system, such as remote inspection.
BIBLIOGRAPHY


Pulfrich C., "Die Stereoskopie im Dienste der isochromen und heterochromen Photometrie," Naturwissenschaften 10, 1922 (English Extracts).


APPENDIX B

MOTOR SELECTION CALCULATIONS.

Prior to the selection of a suitable motor for any of the axes, it is necessary to first ascertain the inertia value of the loading on each axis. This is achieved by considering the distribution of the mass around a particular axis. Details of the inertia calculation for the vergence axes will be presented here, for the other axes the reader is advised to refer to (Asbery 1994), although all of the load inertia and motor selection values will be presented.

B.1 Inertia Calculation for Vergence Axis.

The initial calculation of the inertia for this axis allows for the cameras to be fixed with an external lens. This gives an over-estimate of the actual inertia value as these lenses were not used in the final design. The final design utilised lenses of a much lower mass, which consequently would also have a lower referred inertia. The actual lens used in the final design has dimensions 30 X 20 mm Dia and a mass of 0.01 Kg, which evaluates to a referred inertia of 0.8Kgcm². This is a much lower referred inertia than the actual lens used in the design case, but as the system was designed as an experimental test bed it was necessary to cater for the worse operation case, which would be when larger auto-focus lens were used.

From the schematic drawing contained in Appendix A it is possible to generate a simplified schematic of a vergence axis, shown in figure B.1. From this schematic the inertia calculations can be performed. This is firstly
carried out for each component individually, all inertia then being shifted to act around the axis of rotation using the Parallel Axis Theorem.

**Camera Inertia Calculation.**

The camera is assumed to be cylindrical with a uniform mass distribution.

Mass 0.016Kg.

Dimensions 40mm X 20mm diameter.

\[ I_c = \frac{M}{48} (3d^2 + 4L^2) \]  

(B.1)

Using equation B.1, where \( I_c \) is the inertia of the camera about its vertical centroid, \( d \) is the diameter and \( L \) is the length.

\[ I_c = 0.025 \text{ Kgcm}^2 \]

Figure B.1 Simplified Schematic of Vergence Assembly.
Cable Inertia Effect Calculation.
It is assumed that the cable which imparts loading on the vergence axis is
given by a cylinder with a uniformly distributed mass equal to twice the
camera mass.
Mass = 0.032Kg
Dimensions 50 X 15mm diameter.

Using equation B.1 gives the inertia of the cable about it's own vertical centroid, $I_w$.

$$I_w = 0.067 \text{ Kgcm}^2$$

Lens Inertia Calculation
It is assumed that the lens is cylindrical, with a uniformly distributed mass.
Mass = 0.17Kg.
Dimensions 50 X 50mm diameter.

Again using equation B.1, the inertia of the lens, $I_l$, about it's own vertical centroid can be evaluated, yielding;

$$I_l = 0.73 \text{ Kgcm}^2$$

Mounting Bracket Inertia Calculation.
The camera mounting is assumed to be of rectangular shape and made from
steel. The inertia for a rectangular block, about it's own vertical centroid is
given by equation B.2.

$$I_b = \frac{M}{12}(a^2 + c^2)$$  \hspace{1cm} (B.2)
Appendix B: Motor Selection Calculations.

Mass = 0.14 Kg
Dimensions 35 X 35 X 15mm.

Using equation B.2, the block inertia, \( I_b \) can be evaluated, yielding:

\[
I_b = 0.29 \text{ Kgm}^2
\]

Inertia Referral.
All of the inertia's which have been evaluated, specify the inertia about the components own vertical centroid. Hence the parallel axis theorem, equation B.3, must be applied to refer this inertia to the vergence axis.

\[
I_v = I + mh^2
\]

where \( I \) is the inertia about the object's own centroid and \( h \) is the distance to the new axis of rotation.

The individual inertia's, the referred inertia and the total load inertia about the vergence axis are given in table B.1.

<table>
<thead>
<tr>
<th></th>
<th>Inertia Kgm(^2)</th>
<th>Distance to Verg. Axis</th>
<th>Referred Inertia Kgm(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camera</td>
<td>0.025</td>
<td>0</td>
<td>0.025</td>
</tr>
<tr>
<td>Cable</td>
<td>0.067</td>
<td>4.5</td>
<td>0.67</td>
</tr>
<tr>
<td>Lens</td>
<td>0.73</td>
<td>5</td>
<td>5.73</td>
</tr>
<tr>
<td>Mount</td>
<td>0.29</td>
<td>0</td>
<td>0.29</td>
</tr>
</tbody>
</table>

|        | Total             | 6.715 Kgm\(^2\) |

Table B.1 Vergence Axis Inertial Loading.
Appendix B: Motor Selection Calculations.

B.2 Vergence Axis Motor Selection.

The vergence axes are required to exhibit a maximum angular velocity and acceleration of 400°/sec and 4000°/sec² respectively (see section 4.2.2.1). Considering a Harmonic Drive RH-5 motor and gearbox combination as a possible solution, there are a number of parameters which must be evaluated and compared with the manufacturers specifications, a summary of which are given in table B.2. The following parameters must be compared with the manufacturers specifications. These parameters are general rules for selecting servo motor drives and can be found in many mechatronic design texts (Dorf 1989).

Load Inertia < 3 X Motor and Gearbox Inertia - for adequate performance.
Load Inertia < Motor and Gearbox Inertia for optimum dynamic response.
Acceleration Torque < Repeated Peak Torque.
Load Speed < Rated Output Speed.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>VERGENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor</td>
<td>RH-5 5505</td>
</tr>
<tr>
<td>Max. Output Speed</td>
<td>110 rpm</td>
</tr>
<tr>
<td>Output Torque</td>
<td>0.59 Nm</td>
</tr>
<tr>
<td>Space Envelope</td>
<td>70 X 21 Dia mm</td>
</tr>
<tr>
<td>Repeated Peak Torque</td>
<td>0.59 Nm.</td>
</tr>
<tr>
<td>Gearbox Ratio</td>
<td>80:1</td>
</tr>
<tr>
<td>Reflected Inertia at Gearbox Output</td>
<td>16 Kgcms²</td>
</tr>
<tr>
<td>Mechanical Time Constant</td>
<td>13.3 msec</td>
</tr>
</tbody>
</table>

Table B.2 A Summary of the RH-5 Motor Specification.
Appendix B: Motor Selection Calculations.

Load Inertia Comparison.
By comparison of the load inertia, table B.1, and the motor inertia, table B.2, it can be seen that the load inertia is significantly lower than the referred inertia of the motor and gearbox combination. Therefore the first two requirements are satisfied.

Acceleration Torque Evaluation.
The acceleration torque which is required can be calculated from equation B.4.

\[ T_a = a(I_v + I_m) \] (B.4)

The acceleration torque is evaluated to be;
\[ T_a = 0.15 \text{Nm}. \]

Which can be compared with the value given in table B.2, and can be observed to be much lower than the stated repeated peak torque.

Load Speed.
The maximum output speed required by the vergence axes is 400°/sec and the maximum output speed of the RH-5 motor is 110 rpm or 660°/sec. The required output speed is much lower than the manufacturers stated value, hence the final criteria for motor selection is satisfied.

The suitability of the RH-5 Harmonic Drive motor and gearbox unit has theoretically been evaluated. It has been shown to satisfy all the design requirements.
Appendix B: Motor Selection Calculations.

B.3 Minimum Encoder Resolution Required For The Vergence Axes.

The evaluation of the minimum encoder resolution for the vergence axes is based on determining the angle through which the CCD array, present in the camera, must be moved through in order to move an image from one pixel to the next.

The camera has a 1/2" CCD array which has 681(H) X 582(V) pixels. The horizontal and vertical dimensions of a single pixel can be approximated to be in the order of 10μm. Thus the angle through which the CCD array must be moved to move an image by 10μm on the CCD must be evaluated. It is assumed that the rotation occurs about the focal point of the lens which is being used. In this case it was assumed a lens of 7.5mm focal length was used.

![Diagram](image.png)

Figure B.2 Evaluation of Encoder Resolution.

Figure B.2 shows the encoder resolution is required to be ≤ 0.08°.

For a RH-5 Harmonic Drive motor there are a range of encoders available, the resolutions of which are shown in table B.3.
All the available encoders are able to realise the required resolution, hence a mid range encoder is selected, 300 lines. This offers increased resolution for a comparable cost of the 100 line encoder.
APPENDIX C

STEREOSCOPIC MAGNIFICATION

Stereoscopic magnification, \( M_s \), is given by the equation (Diner and Fender 1993),

\[
M_s = \lim_{\Delta \to 0} \frac{2f_c \tan \phi}{\Delta (l-f_c)}
\]

and, \( \phi = \tan^{-1} \frac{2V}{l} - \tan^{-1} \frac{2(V-\Delta)}{l} \)

Where \( f_c \) is the focal length of the camera lens, \( V \) is the distance to the fronto-parallel plane, \( l \) is the distance between the point of convergence and the lens nodal point.

Due to the dependence of the limit in \( \Delta \), on both the numerator and the denominator. It can be evaluated by differentiating the numerator and denominator with respect to \( \Delta \), Hence,

\[
M_s = \lim_{\Delta \to 0} \frac{\frac{\partial}{\partial \Delta} (2f_c \tan \phi)}{\frac{\partial}{\partial \Delta} (l-f_c)} = \lim_{\Delta \to 0} \frac{\frac{\partial}{\partial \Delta} 2f_c \tan \phi}{(l-f_c)}
\]

Now considering the numerator only,

\[
\frac{\partial}{\partial \Delta} (2f_c \tan \phi) = \frac{\partial}{\partial \phi} (2f_c \tan \phi) \frac{\partial \phi}{\partial \Delta},
\]
Appendix C: Stereoscopic Magnification.

\[
\frac{\partial}{\partial \phi} (2f_0 \tan \phi) = 2f_0 \sec^2 \phi = \frac{2f_0}{\cos^2 \phi}
\]

And

\[
\frac{\partial \phi}{\partial \Delta} = \frac{-1}{1 + \frac{4(V-\Delta)^2}{I^2}} \cdot \frac{-2}{I}
\]

\[
= \frac{2I}{I^2 + 4V^2 - 8\Delta V + 4\Delta^2}
\]

Hence

\[
M_s = \lim_{\Delta \to 0} \frac{4f_0 I}{\cos^2 \phi(1-f_0)(I^2 + 4V^2 - 8\Delta V + 4\Delta^2)}
\]

Now in the limit as \( \Delta \to 0 \),

\[
M_s \to \frac{4f_0 I}{(1-f_0)(I^2 + 4V^2)}
\]

Thus the Stereoscopic Magnification of any dual camera system which is verged onto a point in space may be evaluated for differing values of interocular separation and focal length.

A table of Stereoscopic Magnification values for varying vergence point distances and interocular separation for the telepresence system are presented in table C.1. For this system \( f_0 \) is 7.5mm.
### Appendix C: Stereoscopic Magnification

<table>
<thead>
<tr>
<th>V cm/I</th>
<th>6.50 cm</th>
<th>10 cm</th>
<th>15 cm</th>
<th>20 cm</th>
<th>25 cm</th>
<th>30 cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>1971</td>
<td>3015</td>
<td>4467</td>
<td>5855</td>
<td>7163</td>
<td>8377</td>
</tr>
<tr>
<td>55</td>
<td>1628</td>
<td>2493</td>
<td>3701</td>
<td>4865</td>
<td>5973</td>
<td>7016</td>
</tr>
<tr>
<td>60</td>
<td>1367</td>
<td>2095</td>
<td>3116</td>
<td>4105</td>
<td>5064</td>
<td>5955</td>
</tr>
<tr>
<td>65</td>
<td>1164</td>
<td>1785</td>
<td>2658</td>
<td>3596</td>
<td>4329</td>
<td>5114</td>
</tr>
<tr>
<td>70</td>
<td>1008</td>
<td>1539</td>
<td>2294</td>
<td>3032</td>
<td>3748</td>
<td>4437</td>
</tr>
<tr>
<td>75</td>
<td>874</td>
<td>1341</td>
<td>2000</td>
<td>2646</td>
<td>3276</td>
<td>3884</td>
</tr>
<tr>
<td>80</td>
<td>768</td>
<td>1178</td>
<td>1759</td>
<td>2329</td>
<td>2887</td>
<td>3428</td>
</tr>
<tr>
<td>85</td>
<td>680</td>
<td>1044</td>
<td>1559</td>
<td>2086</td>
<td>2563</td>
<td>3047</td>
</tr>
<tr>
<td>90</td>
<td>606</td>
<td>931</td>
<td>1391</td>
<td>1845</td>
<td>2290</td>
<td>2726</td>
</tr>
<tr>
<td>95</td>
<td>544</td>
<td>835</td>
<td>1249</td>
<td>1657</td>
<td>2058</td>
<td>2452</td>
</tr>
<tr>
<td>100</td>
<td>491</td>
<td>754</td>
<td>1127</td>
<td>1496</td>
<td>1860</td>
<td>2217</td>
</tr>
<tr>
<td>200</td>
<td>122</td>
<td>188</td>
<td>282</td>
<td>375</td>
<td>469</td>
<td>561</td>
</tr>
<tr>
<td>300</td>
<td>54</td>
<td>84</td>
<td>125</td>
<td>167</td>
<td>206</td>
<td>250</td>
</tr>
<tr>
<td>400</td>
<td>31</td>
<td>47</td>
<td>70</td>
<td>94</td>
<td>117</td>
<td>141</td>
</tr>
<tr>
<td>500</td>
<td>20</td>
<td>30</td>
<td>45</td>
<td>60</td>
<td>75</td>
<td>90</td>
</tr>
<tr>
<td>600</td>
<td>14</td>
<td>21</td>
<td>31</td>
<td>42</td>
<td>52</td>
<td>63</td>
</tr>
<tr>
<td>700</td>
<td>10</td>
<td>15</td>
<td>23</td>
<td>31</td>
<td>38</td>
<td>46</td>
</tr>
<tr>
<td>800</td>
<td>8</td>
<td>12</td>
<td>18</td>
<td>23</td>
<td>29</td>
<td>35</td>
</tr>
<tr>
<td>900</td>
<td>6</td>
<td>9</td>
<td>14</td>
<td>19</td>
<td>23</td>
<td>28</td>
</tr>
<tr>
<td>1000</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>15</td>
<td>19</td>
<td>23</td>
</tr>
</tbody>
</table>

Table C.1 Stereoscopic Magnification as a Function of I and V.
APPENDIX D

PRELIMINARY SYSTEM MODEL.

The pan axis linear system model is shown in figure D.1. The parameters used in the model are as follows; \( K_b = 2.8648 \text{ Vsrad}^{-1}, R_a = 2.7 \Omega, K_t = 2.92 \text{ NmA}^{-1}, J_t = 0.021 \text{ Kgm}^2 \). The gain \( K \), shown in the system model simply allows the input and output to be specified in degrees.

The response of the system model to a 45° step input is shown in figure D.2.

![Figure D.1 System Model Step Response.](image-url)
Figure D.2 Pan Axis System Model.
APPENDIX E

FORMULATION OF GOVERNING EQUATIONS.

Any straight line in space can be specified by defining the direction of the line and a point on the line, given by the direction vector \( \mathbf{b} \) and the point \( A \) in figure E.1.

A vector equation of this straight line can be obtained by expressing the position vector \( \mathbf{r} \) of a general point \( P \), located on the line.

Let \( P(r) \) be any point on the line shown in figure E.1, then the vector \( \mathbf{AP} \) is parallel to the direction vector \( \mathbf{b} \) and is therefore equal to \( \lambda \mathbf{b} \) where \( \lambda \) varies according to the position of \( P \).

Hence;

\[
\mathbf{r} = \mathbf{a} + \lambda \mathbf{b}
\]
Appendix E : Formulation of Governing Equations.

which can be reduced for the case of a straight line through the origin, where \( a = 0 \), to,

\[
\mathbf{r} = \lambda \mathbf{b}
\]

This represents the equation of the line of sight vector which can be rewritten in terms of the pan and elevation angles of the telepresence system, shown in figure E.2.

Figure E.2 A Schematic of the Line of Sight Vector.
Therefore,

\[ \mathbf{r}_i = \lambda_i \begin{pmatrix} \cos \phi_i \sin \theta_i \\ \cos \phi_i \cos \theta_i \\ \sin \phi_i \end{pmatrix} \]  

(E.1)

As can be observed from equation E.1, for any given \( \phi_i \) and \( \theta_i \) angles, the value of \( \mathbf{r} \) is related to the coefficient \( \lambda_i \). Consequently, if \( \lambda_i \) is unknown then there are infinite solutions for \( \mathbf{r} \). Therefore in order to determine the position vector \( \mathbf{r}_i \), \( \lambda_i \) must be constrained to a particular value. To perform this in the context of the target tracking experiments demonstrated in this thesis, a value for \( \lambda_i \) can be established as follows.

Consider two points, \( \mathbf{P}_1 \) and \( \mathbf{P}_2 \), on the target plane as shown in figure E.2. Then the line of sight vectors to these two points are given by equations E.2 and E.3 respectively.

\[ \mathbf{P}_1 = \lambda_1 \begin{pmatrix} \cos \phi_1 \sin \theta_1 \\ \cos \phi_1 \cos \theta_1 \\ \sin \phi_1 \end{pmatrix} \]  

(E.2)

\[ \mathbf{P}_2 = \lambda_2 \begin{pmatrix} \cos \phi_2 \sin \theta_2 \\ \cos \phi_2 \cos \theta_2 \\ \sin \phi_2 \end{pmatrix} \]  

(E.3)

In terms of Cartesian co-ordinates the locations of points \( \mathbf{P}_1 \) and \( \mathbf{P}_2 \) can also be written as shown in equations E.4 and E.5.

\[ \mathbf{P}_1 = \begin{pmatrix} \mathbf{P}_{1x} \\ \mathbf{P}_{1y} \\ \mathbf{P}_{1z} \end{pmatrix} \]  

(E.4)

\[ \mathbf{P}_2 = \begin{pmatrix} \mathbf{P}_{2x} \\ \mathbf{P}_{2y} \\ \mathbf{P}_{2z} \end{pmatrix} \]  

(E.5)
Consider a vector $\mathbf{d}$ which joins points $\mathbf{P}_1$ and $\mathbf{P}_2$ as shown in figure E.2, therefore the vector $\mathbf{d}$ must be equal to the vector equation,

$$\mathbf{d} = \mathbf{P}_2 - \mathbf{P}_1 \quad (E.6)$$

or alternatively substituting equations E.4 and E.5 giving equation E.7,

$$\mathbf{d} = \begin{pmatrix} P_{2x} \\ P_{2y} \\ P_{2z} \end{pmatrix} - \begin{pmatrix} P_{1x} \\ P_{1y} \\ P_{1z} \end{pmatrix} \quad (E.7)$$

In order to gain a solution for $\lambda_i$, a number of constraints must be introduced into the system. If points $\mathbf{P}_1$ and $\mathbf{P}_2$ are said to lie on the same vertical and horizontal plane then the following are true,

$$P_{2y} - P_{1y} = 0 \quad (E.8)$$
$$P_{2z} - P_{1z} = 0 \quad (E.9)$$

If these points are further constrained and are said to be separated by a horizontal distance of $\Delta$, then equation E.10 is also true,

$$P_{2x} - P_{1x} = \Delta \quad (E.10)$$

Therefore from equations E.7, E.8, E.9 and E.10 equation E.11 can be formed,

$$\mathbf{d} = \begin{pmatrix} \Delta \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \Delta \\ 0 \\ 0 \end{pmatrix} \quad (E.11)$$
Appendix E: Formulation of Governing Equations.

Now in terms of the vector equations of the line of sight vectors to points $P_1$ and $P_2$, equation E.11 can be rewritten to yield equations E.12, E.13 and E.14,

\[
\begin{align*}
\lambda_2 \cos \phi_2 \sin \theta_2 - \lambda_1 \cos \phi_1 \sin \theta_1 &= \Delta \quad (E.12) \\
\lambda_2 \cos \phi_2 \cos \theta_2 - \lambda_1 \cos \phi_1 \cos \theta_1 &= 0 \quad (E.13) \\
\lambda_2 \sin \phi_2 - \lambda_1 \sin \phi_1 &= 0 \quad (E.14)
\end{align*}
\]

Substitution of equation E.14 into equation E.12 yields after some manipulation equation E.15, where $\Delta$ is the horizontal separation of the two points on the target plane, $\lambda$ is the vector coefficient, $\theta$ and $\phi$ are the pan and elevation angles of the telepresence system respectively.

\[
\frac{1}{\lambda_1} = \frac{1}{\Delta} \left[ \frac{\sin \phi_1 \sin \theta_2 - \cos \phi_1 \sin \theta_1}{\tan \phi_2} \right] \quad (E.15)
\]

And from equations E.2 and E.15 an expression for $P_{1y}$ in terms of the orientation angles, $\phi$ and $\theta$, of the telepresence head and the separation, $\Delta$, between the two points $P_1$ and $P_2$ can be evaluated and is given in equation E.16. This is the distance from the telepresence coordinate frame to the target plane.

\[
P_{1y} = \Delta \frac{\tan \theta_2 \cos \phi_1}{\tan \phi_1 \sin \theta_2 - \tan \theta_1 \tan \phi_2 \cos \phi_1} \quad (E.16)
\]

So, from a given set of 2 points on the target plane which satisfy the constraints of equations E.8, E.9 and E.10, it is possible to calculate the distance to the target plane using the pan and tilt orientations of the telepresence system.
Therefore, for any set of telepresence head orientations it is possible to calculate the Cartesian co-ordinates of the viewpoint point on the target plane, using the distance \( P_y \) determined above. These locations, given by \( T_x \), \( T_y \), \( T_z \) in figure E.3, are the Cartesian co-ordinates of the viewpoint in terms of the telepresence system co-ordinate frame.

\[
\begin{align*}
\theta &= \text{Angel of Pan}, \\
\phi &= \text{Angle of Elevation}.
\end{align*}
\]

**Figure E.3 A Generic Point in Three Dimensional Space.**

Now consider a generic three dimensional point in space, as shown in figure E.4, where the point is located at some known distance \( Y \) and \( \theta \), \( \phi \) are the angles of pan and tilt respectively. The \( X \) and \( Z \) Cartesian co-ordinates can be evaluated using simple trigonometry, yielding,

\[
\begin{align*}
T_x &= Y \tan \theta \\
T_y &= Y
\end{align*}
\]
Appendix E: Formulation of Governing Equations.

\[ T_x = Y \frac{\tan \phi}{\cos \theta} \]

Therefore any viewpoint location on the target plane can be represented as the Cartesian co-ordinate, in terms of the telepresence co-ordinate frame, shown in equation E.17.

\[ P = \left( Y \tan \theta, Y \frac{\tan \phi}{\cos \theta} \right) \]  

(E.17)

This Cartesian co-ordinate representation of the viewpoint location is used in the evaluation of the telepresence to robot co-ordinate transformation.

The relationship between the robot frame of reference and the telepresence co-ordinate frame may be written in terms of a homogeneous transformation, as shown in equation E.18.

\[
\begin{bmatrix}
C_{11} & C_{12} & C_{13} & C_{14} \\
C_{21} & C_{22} & C_{23} & C_{24} \\
C_{31} & C_{32} & C_{33} & C_{34} \\
C_{41} & C_{42} & C_{43} & C_{44}
\end{bmatrix}
\begin{bmatrix}
T_x \\
T_y \\
T_z \\
1
\end{bmatrix}
= \begin{bmatrix}
R_x \\
R_y \\
R_z \\
1
\end{bmatrix}
\]  

(E.18)

where the Cij parameters are functions of the rotations and translations between the two co-ordinate frames, \( T_x, T_y, T_z \) are co-ordinates in the telepresence co-ordinate frame and \( R_x, R_y, R_z \) are the target co-ordinates in the robot co-ordinate frame.

The parameters \( C_{41}, C_{42}, C_{43} \) represent the perspective component of the homogeneous transformation and in this instance can be set to 0 and \( C_{44} \) can set to unity. Thus simplifying the evaluation of the solution set.
Appendix E: Formulation of Governing Equations.

Therefore, for a single point pair, equation E.18 can be rewritten as three equations E.19, E.20 and E.21,

\[
\begin{align*}
R_x &= C_{11} T_x + C_{12} T_y + C_{13} T_z + C_{14} \\
R_y &= C_{21} T_x + C_{22} T_y + C_{23} T_z + C_{24} \\
R_z &= C_{31} T_x + C_{32} T_y + C_{33} T_z + C_{34}
\end{align*}
\]  
(E.19)  
(E.20)  
(E.21)

The equations given in E.19 to E.21 can be further manipulated to yield an equation where the homogeneous transformation parameters form a column matrix of dimensions 12 x 1, as shown in equation E.22.

\[
\begin{bmatrix}
R_x \\
R_y \\
R_z
\end{bmatrix} =
\begin{bmatrix}
T_x & T_y & T_z & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & T_x & T_y & T_z & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & T_x & T_y & T_z & 1
\end{bmatrix}
\begin{bmatrix}
C_{11} \\
C_{12} \\
C_{13} \\
C_{14} \\
C_{21} \\
C_{22} \\
C_{23} \\
C_{24} \\
C_{31} \\
C_{32} \\
C_{33} \\
C_{34}
\end{bmatrix}
\]  
(E.22)

Thus the number of unknowns has been reduced to twelve and equation E.22 defines the relationship between a single telepresence viewpoint and robot/target co-ordinate pair, in terms of the calibration coefficients.

In order to find a solution for the homogeneous transformation coefficients, given in equation E.22, it is necessary to find the inverse of the 3 x 12 telepresence co-ordinate matrix. This can be achieved by adopting a pseudo inverse and least squares approach. This approach requires the system of equations to be over determined in order for the telepresence co-ordinate matrix to exhibit full column rank. The matrix to be solved can be formed.
from the consideration of a number of telepresence/robot co-ordinate pairs. This requires a minimum of 5 pairs of co-ordinates which are not all coplanar.

In practice a larger number of co-ordinate pairs is used to improve the accuracy of the least squares fitting technique. In this thesis an array of 12 points, located on two distinct planes, are used to perform the calculation for the target trajectory tracking experiments. A detailed description of the techniques available to solve a set of over determined linear equations, such as these, can be found in many linear algebra texts (Chapra and Canade 1989).
This appendix investigates the effect of modelling errors on the results of the target tracking experiments performed in Chapter 7 of this thesis.

Figure F.1 Modelling Offset.
In the experiments the model of the telepresence system is approximated such that the vertical rotation is assumed to take place around the same axis as the horizontal rotation. This introduces some errors into the calculation of the viewpoint location, the position which is calculated is shown by $P_2$ in figure F.1, the $Z$ co-ordinate of which is described by equation F.1. This assumes that the axes are located at $O$ in figure F.1.

\[ P_{2z} = \frac{Y \tan \phi}{\cos \theta} \quad \text{(F.1)} \]

where $Y$ is the distance to the target plane, $\theta$ and $\phi$ are the pan and tilt angles of the telepresence head respectively.

If the offset of the two axes, shown by $\gamma$ in figure F.1, is taken into account when evaluating the viewpoint location, the $Z$ co-ordinate of point $P_1$ is calculated as shown in equation F.2.

\[ P_{1z} = \frac{Y - \gamma}{\cos \theta} \tan \phi \quad \text{(F.2)} \]

The error in the calculation of the vertical position of the viewpoint, $\delta$, can be established, and is shown in equation F.3.

\[ \delta = P_2 - P_1 = \frac{Y \tan \phi}{\cos \theta} - \frac{(Y - \gamma)}{\cos \theta} \tan \phi \]

\[ \delta = \frac{\gamma}{\cos \theta} \tan \phi \quad \text{(F.3)} \]
Therefore, the error is shown to be a function of $\gamma$, $\theta$ and $\phi$, where $\gamma$ is a constant, determined by the design of the telepresence system and is equal to a magnitude of 50mm.

During the target tracking experiments, Section 7.3, the telepresence system exhibited the following ranges of motion.

$$-16^\circ \leq \theta \leq 31^\circ$$
$$-31^\circ \leq \phi \leq 17^\circ$$

The error, $\delta$, due to the approximation of the location of the elevation axis is plotted on figure F.2 for the ranges of motion experienced during the experiments. Observation of figure F.2, shows that the relationship is non-linear and the error is greater at the extremes of the workspace. This error approximates to value of 58 mm at the edge of the workspace.

![Figure F.2 Modelling Error Over the Workspace.](image)

This shows that there is a degree of error involved in the evaluation of the viewpoint location due to modelling assumptions However it is not the
Appendix F: Modelling Error.

absolute position of the viewpoint which has been of interest, but a comparison between a number of experiments which are all subjected to the same approximation. Consequently all the results will be affected by the same amount.