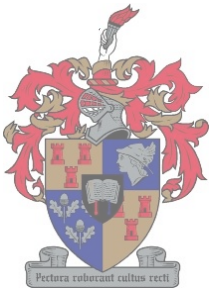
A COMPARATIVE STUDY OF

TARGET ACQUISITION PERFORMANCE OF

TELEOPERATED SENSORS

AND DIRECT HUMAN VISION



BY

G T VILJOEN

Dissertation presented for the degree of

Doctor of Philosophy

at the University of Stellenbosch

November 1999

Dissertation approved for the degree of Doctor in Philosophy at the University of Stellenbosch.

Promotor: Prof. G W Milne

DECLARATION

I, the undersigned hereby declare that the work contained in this dissertation is my own original work and that I have not previously in its entirety, or in part, submitted it at any university for a degree.

G T Viljoen ..

Date 17 NOVEMBER 1999.

ACKNOWLEDGEMENTS

I wish to extend my thanks to my study leader, Prof. Garth Milne for his guidance, help and persistence throughout my studies. To Wollie Wolmarans from Armscor, whose funding helped to make this thesis possible, thanks for your support and your confidence in this project. Finally I would like to thank my wife and loving companion Helene for her unwavering support through the years of hardship, loneliness and sacrifice, without you this thesis would never have been completed.

ABSTRACT

This dissertation shows that by improving teleoperator sensor design, similar performance to direct human vision can be achieved on a target acquisition task. Current technology teleoperated sensors give inferior target acquisition performance compared to direct human vision, as demonstrated by a set of target acquisition experiments with a simulated television sighting system versus direct human vision on the same natural target and background scenes. The main reasons for the former's poor performance were ascribed to tunnel vision and the dynamic presentation on the television display that reduced the visual acuity of the subjects. It was further shown that the human visual system could be adequately stimulated over a 60° field of view with only two video channels - a low angular resolution, wide-field for peripheral vision and a high angular resolution, narrow-field for foveal vision. By displaying the wide-field and narrow-field information into separate eyes the visual cortex could integrate the two pictures. Experimental proof was provided with a simple cardboard tester, followed by an electro-optical tester. It was then postulated that an eye-slaved foveal inset display system would give similar target acquisition performance to unaided human vision. An eye-slaved foveal inset display system was built and tested on the same target scene used for the direct vision and the television sight experiments. Waldman's target acquisition theory was extended to include saccadic movements using Stark's scanpath theory to model fixation points in the given scene. The mathematical model was calibrated with the experimental data. The validated theory was used to demonstrate that target acquisition performance close to direct human vision could be achieved if high definition displays were used. The applications where the foveal inset display system provided benefits were also defined. Further work is recommended on higher resolution, colour, and stereoscopic designs.

ABSTRAK

Hierdie verhandeling toon dat deur teleoperator sensor ontwerp te verbeter, teikenopsporingsprestasie soortgelyk aan die van direkte visie verkry kan word. Huidige tegnologie teleoperator sensors gee swakker teikenopsporing prestasie as direkte visie en dit was gedemonstreer deur 'n stel teikenopsporingseksperimente met onderskeidelik 'n gesimuleerde televisie visierstelsel, en direkte menslike visie van dieselfde natuurlike teiken en agtergrondstoneel. Die hoofrede vir die swak prestasie van die televisie visierstelsel word toegeskryf aan tonnelvisie en die dinamiese voorstelling op die televisiebeeld wat lei tot verlaagte oogskerpte. Daarna is getoon dat die menslike visuele stelsel voldoende gestimuleer kon word oor 'n 60 grade vertoonveldwydte met net twee video kanale – een lae hoekresolusie, wyeveld beeld vir periferale visie, en 'n tweede, hoë hoekresolusie nouveld beeld vir foveale visie. Verder is die wyeveld en nouveld inligting apart in die proefpersoon se oë vertoon, en sy visuele korteks het die twee prentjies geintegreer. Eksperimentele bewys vir laasgenoemde is verskaf deur 'n direkte visie kartontoetser, opgevolg deur 'n elektro-optiese toetser. Daar is gepostuleer dat 'n ooggeslaafde, foveale-inset vertoonstelsel soortgelyke teikenopsporing prestasie sal gee as direkte visie. Laasgenoemde stelsel is gebou en getoets op dieselfde tonele van die vorige eksperimente. Waldman se teikenopsporingsmodel is uitgebrei om sakadiese oogbewegings in te sluit gebasseer op Stark se skandeerpad teorie om die oogfiksasies in die toneel te modelleer. Die wiskundige model is gekalibreer met die eksperimentele data. Die gevalideerde model is gebruik om te demonstreer dat soortgelyke teikenopsporingsprestasie aan direkte visie verkry kon word, as hoë-definisie televisievertone gebruik word. Die toepassings waar die foveale inset vertoonstelsel voordele bied was ook gedefinieer. Verdere werk op hoër resolusie, kleur en stereoskopiese ontwerpe word aanbeveel.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Az	Azimuth (In the horizontal plane)
CCD	Charge Coupled Device
CRT -	Cathode Ray Tube
dpi	Dots per Inch
DSP	Digital Signal Processor
DVT	Direct Vision Test
El	Elevation (In the vertical plane)
ESFID	Eye-slaved foveal inset display
ESHDS	Eye-slaved helmet display system, the same as ESFID
FOV	Field of View
H	Horizontal
HDTV	High Definition Television
HMD	Helmet mounted display
LOS	Line of sight
MIT	Massachusetts Institute of Technology
NF	Narrow field
NOSC	Naval Ocean Systems Centre
PAL	Phase Alternating Line Television standard
Pacq	Probability of target acquisition
Rx	Receiver
SFP	Split-eye foveal/peripheral
Std.Dev.	Standard Deviation
STV	Surrogate Teleoperated Vehicle
TDS	Terminal Display System
TV	Television
Tx	Transmitter
UAV	Unmanned Air Vehicle
V	Vertical
WF	Wide Field

CHAPTER 1 INTRODUCTION

1.1. BACKGROUND

This dissertation addresses the problem of remote vision. The ideal remote vision system collects all the visual wave front information that the human visual system can sense at a remote sight, transmits the information to a remote operator, and reconstructs the information in real time on a 360 degree, all-round display system. The operator is not able to visually discern a difference between being at the remote site or not. Tachi [1] envisaged this idealised remote vision concept as illustrated in figure 1.1. The work in this thesis is aimed at those applications where this idealised performance is required, but it should be noted that the opportunity for geometric, dynamic or symbolic enhancements of the user-interface makes it possible to provide interfaces that incompletely reproduce the physical aspects of the remote site.

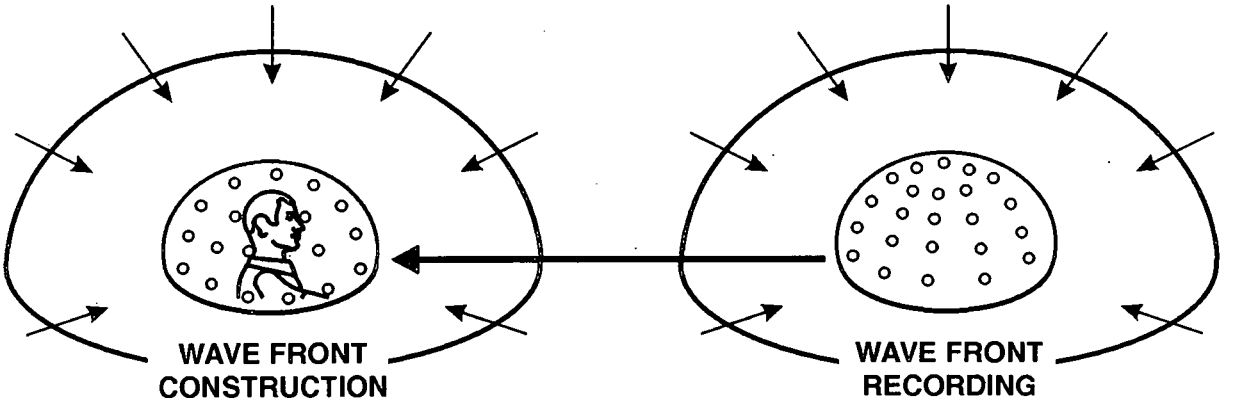


FIGURE 1.1 AN IDEAL REMOTE VISION SYSTEM

Human vision has a peak resolution of between 150-300 μrad (0.5-1 arcmin)[2] depending on a number of factors, including the luminance level of the target being viewed. An omni-visual display as shown in figure 1.1, with a 300μrad (1 arcmin) resolution, will require more than 440 million pixels that have to be updated at more than 50 Hz [3]. True colour vision requires a 24bits/pixel colour display [4], which will bring the required data rate to around 65 gigabyte per second. The remote vision challenge is to achieve these idealised specifications, or their equivalent. Current and near future technology recorders, communications systems and displays cannot

even closely match the resolution and bandwidth requirements [5]. In this thesis, an eye-slaved foveal inset display design is presented that approximates the above-mentioned ideal information display performance. The design uses existing technology building blocks that exploit some of the physiological aspects of the human visual system to meet the display requirements.

Remote controlled vision systems can be regarded as a sub-class of teleoperators, according to Sheridan [6]. The word teleoperator comes from "Tele"-at a distance and "operator" that which operates. To operate can mean to transform the physical environment by an energy exchange or it can mean to transform information from one configuration to another. For example, visual wave front data is transformed to traces on a display CRT. In this thesis the teleoperator is a remote-controlled, visual target acquisition system, or a telesensor system.

Teleoperation is used under the following circumstances:

a) Where tasks are unpredictable and human perception, planning and control are required – i.e. man-in-the-loop operation. Witkowski shows that teleoperators are used in those industries where work is being performed that requires a level of manipulative and perceptual skills far beyond that presently obtainable with industrial robots [7].

b) Where work sites are hazardous to human health or survival. Outer space [8], deep-ocean [9], nuclear and toxic environments [10], mines, construction sites, fire-fighting [11], police and military operations [12] are all hazardous to the human operators involved. Teleoperated machines remove the operator from these hazardous environments while enabling the operators to still execute their tasks from a remote location.

c) Teleoperators are also used where extremely small and dextrous operations are required. An example is telerobotic aiding of laparoscopic surgery [13].

Telesensors evolved from direct viewing devices such as glass windows, through optical and periscope viewing devices to television, where electronic imaging, transmission and display is used. A modern telesensor system consists of a television sensor, a communications link and a display system to the operator. Remote viewing systems can be classified as follows:

a) Direct viewing through air and special windows.

b) Viewing through optical devices that include binoculars, fibre-scopes, episcopes, periscopes and other optical sights.

c) A joystick controlled pan and tilt camera with a panel mounted television display, also called a terminal display system in this thesis.

d) Head slaved cameras with either monoscopic or stereoscopic helmet mounted displays.

Current conventional television technology gives between 480-1080 lines of resolution, depending on the format ranging from NTSC to HDTV[14]. A television display that matches peak human visual acuity spans a field of view of only 7-10 degrees diagonally. A remote controlled visual sensor, which gives the same resolution as human foveal vision, will thus be limited to an extremely narrow field of view, giving rise to tunnel vision. Military sights use wide, intermediate and narrow field of view designs between which the operator can switch to help reduce this tunnel vision effect, but it takes a long time to find and position each object of interest in the centre of the display. A pan and tilt camera platform that is slaved to a helmet mounted display system can expand the field of regard considerably. The helmet-mounted display will have either a wide field, low-resolution display (such as used on current Virtual Reality displays [5]) or a narrow field, high-resolution display. Current display technology does not give the resolution required for full remote human visual acuity over the total field of view and regard of human vision.

This dissertation is aimed at the application of teleoperator technology on a long range (>100 m) target acquisition task, where the viewing is not extended and stereopsis has a small effect. The target acquisition performance of the teleoperator is then compared to the performance of direct human vision, given the same task. The control experiment is the acquisition performance of direct human vision. An existing technology, terminal display system with joystick control of a pan and tilt camera system is used as the control subject for a telesensor system. The loss in acquisition performance that the terminal display system design incurs is measured relative to direct human vision.

A split-view display design was developed that regains the loss in target acquisition performance of the terminal display telesensor design, and achieves comparable performance to direct human vision. The contribution of this thesis lies in explaining when and how the split view sensor improves performance above a single view sensor. The research required significant experimentation to generate results that were incorporated in a mathematical model of the acquisition process from which generalisations could be made in terms of the design parameters.

1.2. THESIS STRUCTURE

The first chapter defines the problem that is going to be studied, as well as the thesis statement and the aims of the research.

Chapter 2 surveys existing literature, defines the terms used in this thesis, after which the current status of the research in the field is discussed and this thesis' contribution is put in perspective to that research. Trends and gaps in the field of study are defined. A mathematical model of the problem is developed, based on the literature.

Chapter 3, 'Method' describes the research method that was used. The advocated solution to the problem statement is then developed, based on a number of hypotheses, which had to be tested. The experimental program is described including the experimental protocols and apparatus.

Chapter 4 describes each of the experiments and presents the results of each of the experiments defined in the method section. The results of the experiments are also compared to each other.

The results are discussed and generalised in Chapter 5. The implication and validity of the experimental data is discussed and analysed. The generalisation is done with a mathematical model validated by experimental results. The limits to the generalisation are determined, and the thesis statement is validated. The areas where the advocated system can provide benefits are defined.

In Chapter 6, 'Conclusions', the research results are summarised, the research contribution is summarised and conclusions are drawn. Recommendations are made for future research.

The appendices give the detailed mathematical models and a detailed description of the ESFID hardware.

1.3. PROBLEM DEFINITION

1.3.1. Problem Description

The problem that is studied is to determine the differences in target acquisition performance between direct human vision and teleoperated sensors. This study is constrained to long-range targets and in monochrome only. Two types of teleoperated target acquisition systems are studied. The first system is a conventional technology, terminal display system with joystick control. The second system is an eye-slaved foveal inset display system. The main differences between these two teleoperated systems are firstly the display (eye-slaved vs. a stationary terminal) and secondly the control (joystick vs. eye movement itself). The problem is shown graphically in figure 1.2. The teleoperator system structure comes from Niemeyer [15].

This dissertation concentrates on the target acquisition subsystem level. A target acquisition theory developed on the subsystem level can be applied on a range of teleoperated target acquisition systems.

The teleoperated target acquisition problem occurs in many potential fields of interest, for example:

a) The Military

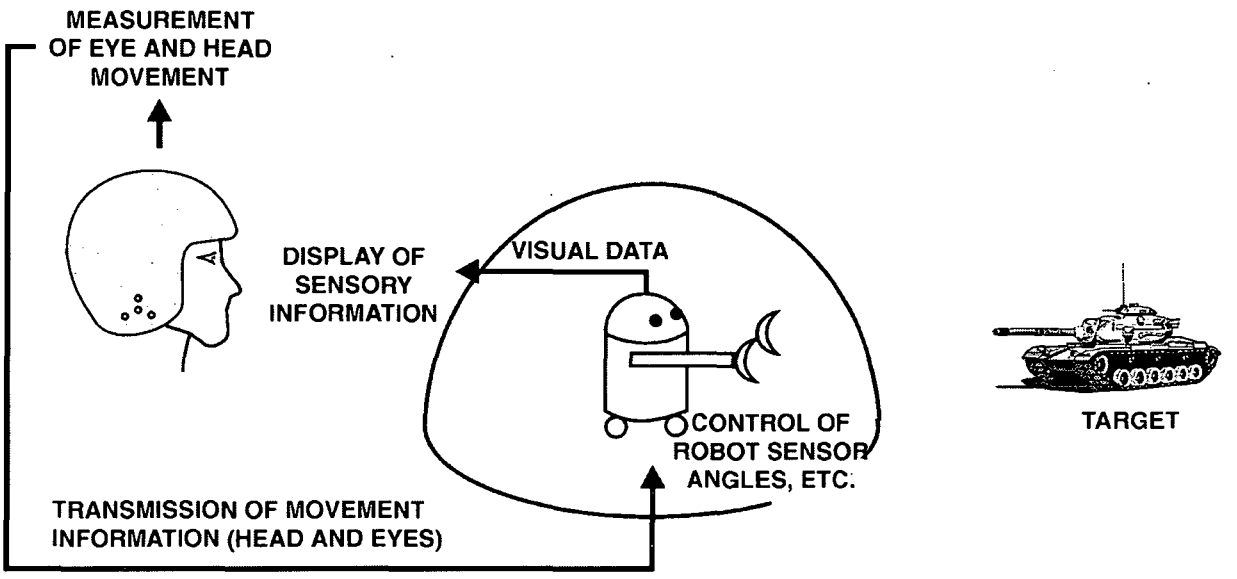
Teleoperators can play an important role in acquiring, tracking and engaging enemy targets without exposing the operator to direct fire. The teleoperator can be mounted on a remotely piloted ground or air vehicle [16], an elevated mast [17] or at a remote site [18]. This technology will lead to a higher survivability of own forces.

b) Policing, border patrol and peace keeping

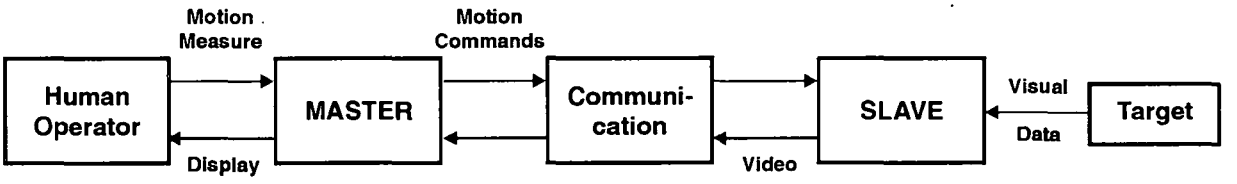
In all these cases an unmanned air vehicle with a teleoperated target acquisition system can be used to acquire and track suspects from the air [19].

The rest of this chapter defines the problem further in terms of the task, the human operator, the teleoperator and the performance measure definition.

a) Teleoperated system



b) Teleoperator Block diagram



c) Direct Vision

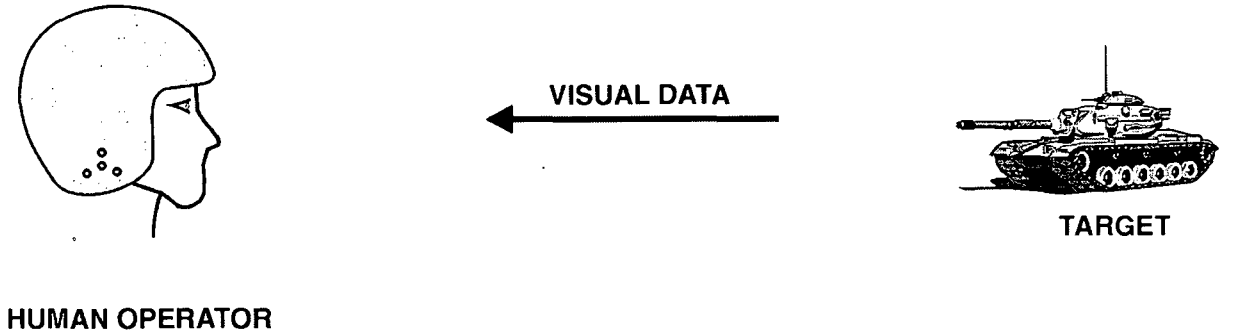


FIGURE 1.2 TELEOPERATED VS. HUMAN ONLY TARGET ACQUISITION: A) TELEOPERATOR SYSTEM DIAGRAM, B) TELEOPERATOR BLOCK DIAGRAM, C) DIRECT VISION

1.3.2. Task definition

The target is the object of interest and it can be a person, a vehicle or a man-made body of interest. The target may be in the presence of a large number of other similar looking objects that clutter the environment and complicate the acquisition aspects. A single military target, as defined by Johnson[20] was used, and the T-72 main battle tank was selected as the target. Only one target was included in the environment at a time.

Bailey [21] described target acquisition as the collective name for the activities search, detection, classification, recognition and initiation of target tracking.

a) Search is defined as the scanning of an area of interest with a target acquisition system.

b) Detection is defined as the sensing of an object of interest.

c) Classification assigns the object of interest to a class of objects. For example the object of interest may be assigned into the class vehicle and not shrub.

d) Recognition is defined as the determination of the type of object e.g. the vehicle is a light truck. Recognition requires a good target outline, and is very dependent on the resolution of the acquisition system.

e) Initiation of target tracking is aligning the acquisition system so that the target is in the centre of the field of view.

The targets in this thesis are restricted to long range, also known as far-field targets [22]. Far-field imaging implies that no stereopsis is involved, and that targets are at a distance of more than a 100m [23].

1.3.3. Human operator definition

For the purpose of this thesis the human visual system is represented by a number of subjects, each having 20/20 vision with corrective optical aids such as glasses or contact lenses. The human operator is stationary in the target environment.

1.3.4. Teleoperator definition

Teleoperation systems can be divided into master and slave systems. The master is the local man machine interface and commands the remote slave system, in which the sensors are located. This classification is also shown in figure 1.2.

Two teleoperator designs are defined for this thesis:

a) A joystick controlled teleoperated pan and tilt video camera sensor platform with a stationary terminal display system. This configuration is the basis of the approach followed by Dr. Antal Bejczy of the Jet Propulsion Laboratory for his advanced teleoperator interface [24]. A large number of electro-optic military sight designs also follow this approach. This design is called the terminal display telesensor and it serves as the control of the teleoperation system design experiment.

b) A head controlled, remotely operated pan and tilt video camera sensor platform with an eye-slaved foveal inset display system. The principle is described by Tachi et al. [1]. Kim and Stark described a helmet-mounted display for telerobotics [25]. The one eye's image is steered by the head angle and the other eye's image is steered by the sum of the operator's head and eye angle.

No manipulation or locomotion is included. The operator is stationary in the target environment. Other teleoperator design options, which have not been included in this thesis, include the following:

- Sight line stabilisation.
- Zoom optics.
- Colour.
- A separate, high resolution, target detection system aid.
- Automatic target detection algorithms and cueing.

Due to the complexity and cost of colour, only monochrome sensors are studied in this thesis. In target detection range experiments carried out by Miller [26], it was found that monochrome systems gave 90% of the target detection range performance of colour systems. Also, the colour resolution of the eye is lower than monochrome as the primary colours map to different cone sensors on the retina, whereas monochrome light stimulates all the sensors. The effect of colour is thus a future research topic, but is expected to be small.

1.3.5. Performance measure definition

To measure the performance of the human operator and the teleoperator system, it is important to have a single figure of merit to quantify performance.

Acquisition systems usually use the probability of correct target acquisition (P_{acq}) as their performance measure. Bailey [21] defines target acquisition probability as the product of visibility, detection and recognition probabilities:

$$P_{acq} = P_{(rec|los\&det)} \times P_{(det|los)} \times P_{(los)} \tag{1.1}$$

Where:

$P_{(los)}$ is the probability that target is in the field of view and visible due to topographical screening, vegetative and or structural screening and atmospheric transmission,

$P_{(det|los)}$ is the probability of detecting the target given that it is inside the field of view of the acquisition system, given that a line of sight exists, and

$P_{(rec|los\&det)}$ is the probability of recognising the target, given line of sight and detection.

It is important to realise that these probabilities are dependent on the search time allocated. Another important observation is that the performance measurement is a probability, which means a large number of observations will be required, for high confidence experimental results.

1.4. THESIS DEVELOPMENT

1.4.1. Thesis Statement

The thesis statement, which forms the central theme of this dissertation, is the following:

It is possible to design a teleoperated sensor using present technology to achieve close to the same performance as direct human vision on a single military target acquisition task.

Analysis in this thesis focuses on the single military target to relate to available literature and studies on military target acquisition. For instance, Johnson's [20] target recognition criteria was developed for single military targets ranging in size from a standing man to a main battle tank. This is also the reason why a T72 main battle tank is used as the target for the target acquisition research in this dissertation. The research that was conducted for this dissertation was aimed at determining to what extent the thesis statement is true, what design would be required for the teleoperator to realise similar performance, and under which circumstances it would be true (i.e. the limits to the generalisation). An eye-slaved foveal inset display concept was developed to test the thesis statement. A typical application would be that of "see-through" armour, where the teleoperated sensor allows an operator inside an armoured vehicle to have a full panoramic display of the scene outside the vehicle, while the operator enjoys full armour protection.

1.4.2. The Eye-slaved Foveal Inset Display Concept

The eye-slaved foveal inset display (ESFID) concept is based on the exploitation of a number of physiological phenomena of the human visual system:

a) Instantaneous Field of View

The first phenomenon used is the limited instantaneous field of view of human vision. The maximum field of view is shown in figure 1.3 [27]. The ideal display's field of view requirements (as shown in figure 1.1) can thus be decreased drastically if the display tracks head motion, such as with a helmet mounted display system with the telesensor being controlled by the operator's head movements.

b) Field of fixation

During tasks such as picture scanning and recognition, the human eye moves in a series of rapid movements, called saccades, inter-spaced with fixations on the area being searched. This phenomenon has been described by Noton and Stark [28]. For an object to be detected, the eyes have to fixate on it. The field of fixation of the human visual system is much smaller than the instantaneous field of view and is given by Ogle [27] to be around 90 degrees horizontal and 50 degrees vertical as shown in figure 1.4. The field of fixation is an order smaller than the idealised display of figure 1.1, and it is the maximum display area required for a helmet mounted target acquisition telesensor system.

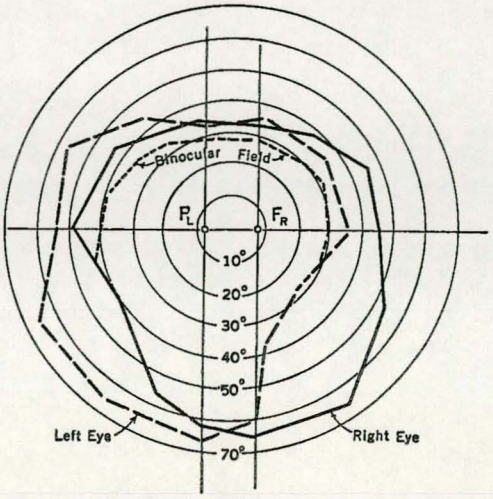


FIGURE 1.3 MAXIMUM INSTANTANEOUS FIELD OF VIEW

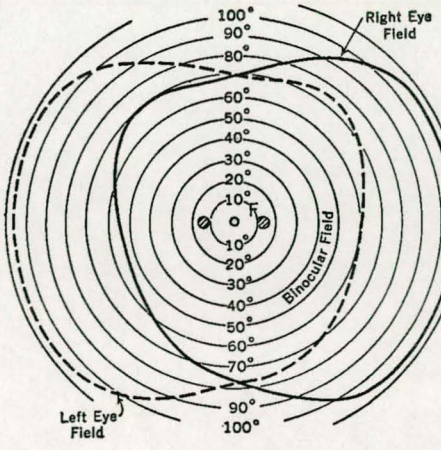


FIGURE 1.4 MAXIMUM FIELD OF FIXATION

c) Visual acuity decreases towards the periphery

Opticians measure the eye's ability to discriminate fine objects or the details of objects in terms of acuity [29]. Acuity is usually expressed as a ratio (x/y, termed Snellen acuity) which indicates that the tested person can barely read at a distance of "y" feet what a normal person can read at "x" feet. "Normal" or 20/20 vision is the ability to resolve a target detail of 1 min of arc (290.89 µrad) at 20 feet. Decimal acuity is x divided by y, and a normal person's visual acuity is widely taken as 1.0. The eye's visual acuity peaks on the centre or foveal part of the eye and decreases rapidly towards the periphery. The isopters, or curves of equal visual acuity, are shown in figure 1.5 [27]. If the display is designed to cover the foveal part of the eye with a resolution of 150 µrad (0.5 arcmin) per pixel, or 300 µrad (1 arcmin) per line pair, it will match the limit of normal visual acuity of a typical eye [2]. Such a foveal vision display would have a vertical field-of-view of 5.0° for a 625 line PAL video system (which provides 577 active lines). The horizontal field-of-view would then be 6.7° if a 4:3 aspect ratio is used, to give a diagonal field of view of 8.4°. If a peripheral vision display with a vertical field-of-view of 48° is combined into the foveal display, a 1.3 mrad/pixel (4.47 arcmin/pixel) resolution will be achieved which equals the eye's resolution at an off-axis angle of around 15°. This display configuration is shown in figure 1.6. This combined foveal/peripheral display gives all the resolution the human eye is capable of over the foveal region of the eye. Between 3° to 15° off-axis the display gives less resolution than the human eye is capable of, but beyond 15° the display gives more data than the eye can use.

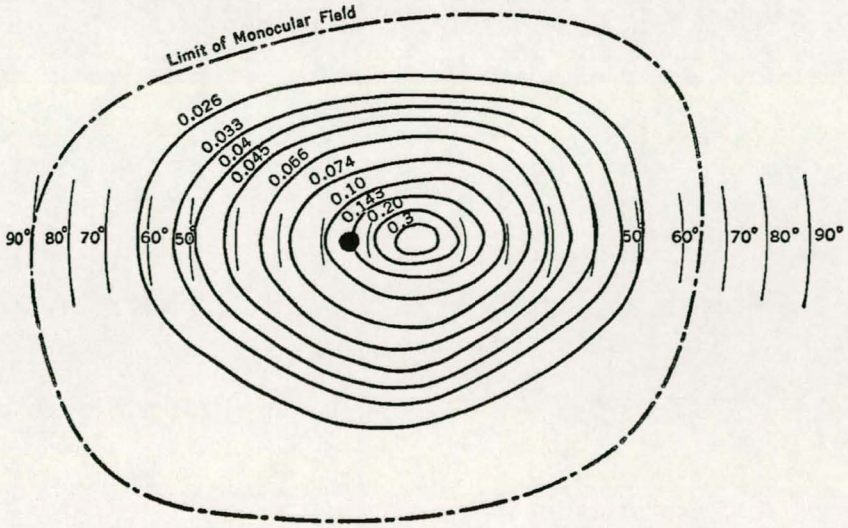


FIGURE 1.5 ISOPTERS OF RELATIVE (TO PEAK) VISUAL ACUITY (I.E. 0.3 ⇒ 70% LOSS IN RESOLUTION) [27]

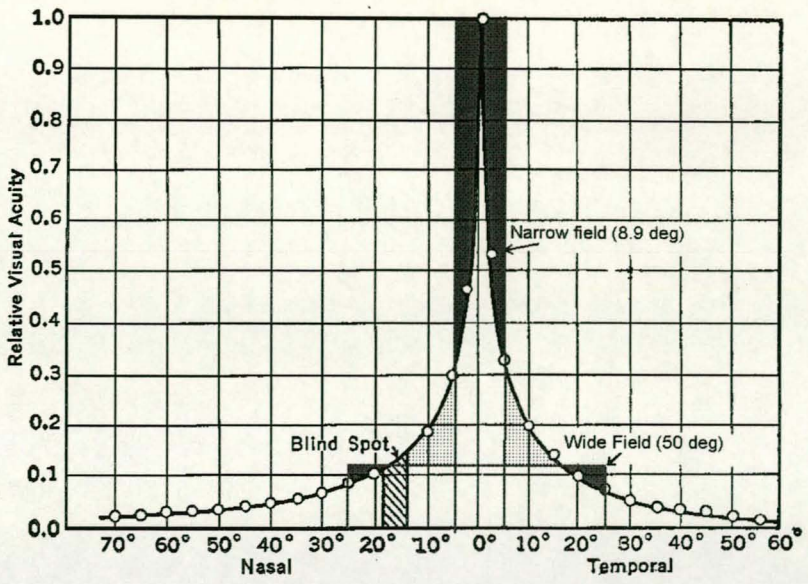


FIGURE 1.6 EYE ACUITY [27] SHOWN WITH FOVEAL AND PERIPHERAL PAL TV DISPLAYS

The following analysis does a more rigorous matching between the eye and the monitor. O.H. Schade [30] determined that the equivalent rectangular pass-band of an optical system is given by:

$$N_e = \int_0^\infty (MTF(\lambda))^2 d\lambda \tag{1.2}$$

Where MTF(λ) is the modulation transfer function of the optical system as a function of the spatial frequency, λ. It is postulated that the eye and monitor will be adequately matched if the equivalent pass-band of the eye and monitor in cascade differs by less than 20% with the eye on its own, i.e. the eye to monitor pass band ratio must be greater than 0.8, or

$$\frac{\int_0^\infty (MTF_{monitor} MTF_{eye})^2 d\lambda}{\int_0^\infty (MTF_{eye})^2 d\lambda} \geq 0.8 \tag{1.3}$$

Figure 1.7 gives the MTF of a PAL TV monitor with 577 active lines (N_{tv}) over a 6.6° vertical field of view (which is the FOV used later in the thesis). A Gaussian spot size is assumed whose MTF is given by equation (1.4) according to the FLIR 92 model as described by Holst [31] with *Zoom* ratio of unity.

$$MTF_{monitor} = e^{-2\left(\pi \sigma_{mon}\left(\frac{f_{raster}}{Zoom}\right)\right)^2} \tag{1.4}$$

and

$$\sigma_{mon} = \frac{1}{2.31 N_{tv}} \tag{1.5}$$

The MTF of the eye in figure 1.7 comes from measurements made by Westheimer [32]. His work was followed up by Williams in 1994 [2, p32] which shows that individual variations between eyes are in the order of 20%. This variation supports the postulate of equation (1.3). The eye to monitor pass band ratio for a 6.6deg vertical FOV PAL TV monitor is equal to 0.87, well within equation (1.3)'s criterion.

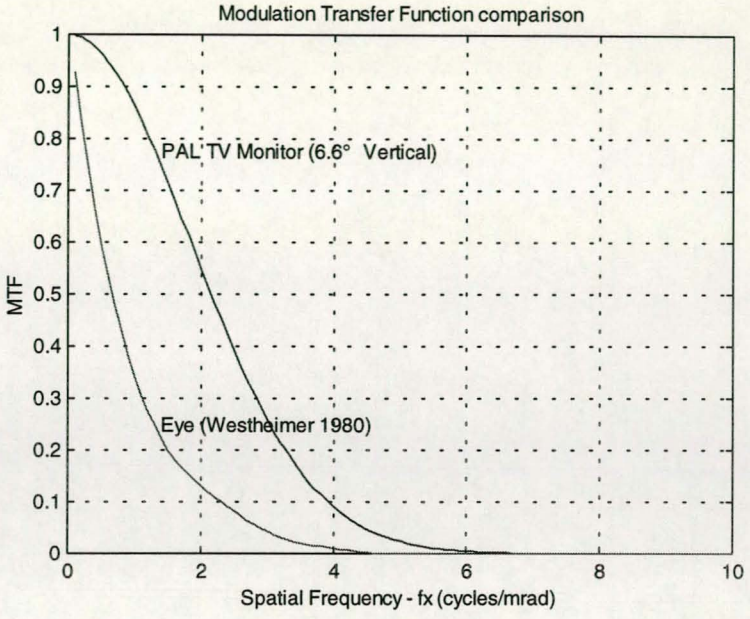


FIGURE 1.7 MODULATION TRANSFER FUNCTION OF EYE AND PAL TV MONITOR

The advent of high definition TV with its high-resolution displays will enhance the ESFID concept to the limit of the human visual system. With the HDTV's maximum resolution format, which gives 1080 lines and a 16:9 aspect ratio [33], the narrow field display will cover 9.28°(V) x 16.5°(H) if the pixel resolution is held at 0.15 μrad/pixel (0.5 arcmin/pixel). At the limit of the HDTV narrow field display, the eye's relative acuity is down to about 0.2 or a resolution of 0.75 mrad/pixel (2.58 arcmin/pixel shown on Figure 1.5). If the wide-field display has the same angular resolution, a field of view of 46.4°(V) by 82.5°(H) will be realised, which covers most of the field of fixation of the human eye (Figure 1.4). A HDTV based display concept should then give human-eye resolution over most of the field of fixation.

d) Split Foveal/Peripheral Fusion

A typical display design that will enable the operator to see both a narrow field and a wide field display in each eye is shown in figure5.9. The shaded area shown is glass. This display will be an extremely heavy design that is quite impractical for head mounted applications. It will also be difficult to achieve wide field of views with this type of display.

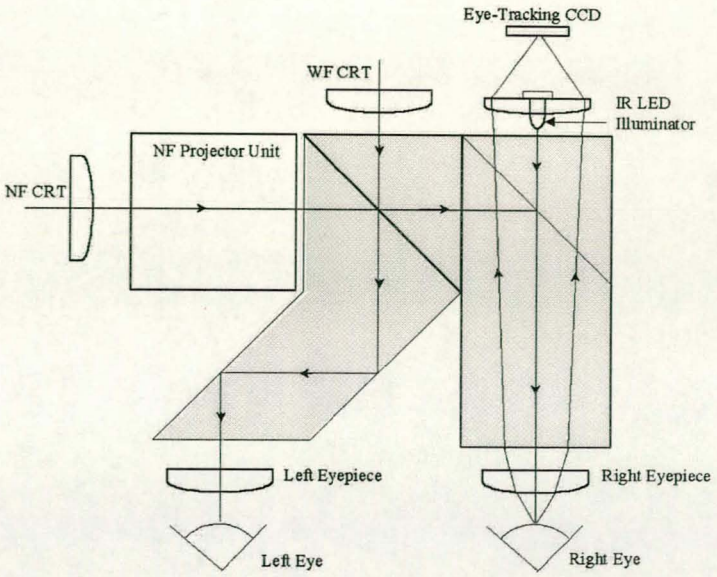


FIGURE 1.8 COMBINED WIDE AND NARROW FIELD STEREO DISPLAY

Early experiments showed that the wide and narrow field displays could be projected into separate eyes and that the human brain could fuse the two displays into one picture [34]. The separation of the wide and narrow field displays was a crucial step in implementing a much simpler display system design than the combined display of figure 1.8, and it paved the way for a practical helmet mounted display.

The split foveal/peripheral fusion phenomenon can be easily demonstrated with a simple test, described in Chapter 4.1. A cardboard sheet is created with a large hole to simulate a wide-field display for the one eye and a small hole in front of the other eye to simulate a narrow field display. The wide-field display's foveal area is covered by a piece of cardboard of the same diameter as the narrow field display. The operator looks at a distant target, (far enough to avoid parallax being a problem) and the foveal information of the one eye will be found to fuse with the peripheral information of the other eye - assumedly in the visual pathway of the brain of the operator.

It is hypothesised that the split foveal/peripheral fusion capability has evolved in the human visual system to compensate for the condition where if one eye suffers a loss of accommodation, the visual system still uses the peripheral information from that eye. Patients suffering from strabismus (or squint-eye) also adapt by ignoring the foveal vision of one eye.

e) Eye movements and suppression of visual function

The final physiological phenomenon, which the ESFID design exploits, is the eye's movement control. During tasks such as picture scanning and recognition, the human eye moves in a series of rapid movements, called saccades, inter-spaced with fixations on the area being searched [28]. Typical results usually show a saccadic suppression in that the detection threshold of the eye drops from 100% about 100 ms prior to the saccade to near 0% shortly after the saccade begins, and climbs gradually to 100% about 150 ms after the completion of the saccade [35]. This suppression period allows time for an opto-electronic system to measure the eye's angle and to command a display to follow the eye's fixations.

1.4.3. Development of the ESFID

The split foveal/peripheral eye-slaved foveal inset display system concept described above was implemented with commercial, off-the shelf equipment and is shown on figure 1.9. The development is described in more detail in chapter 4 and Appendix C.

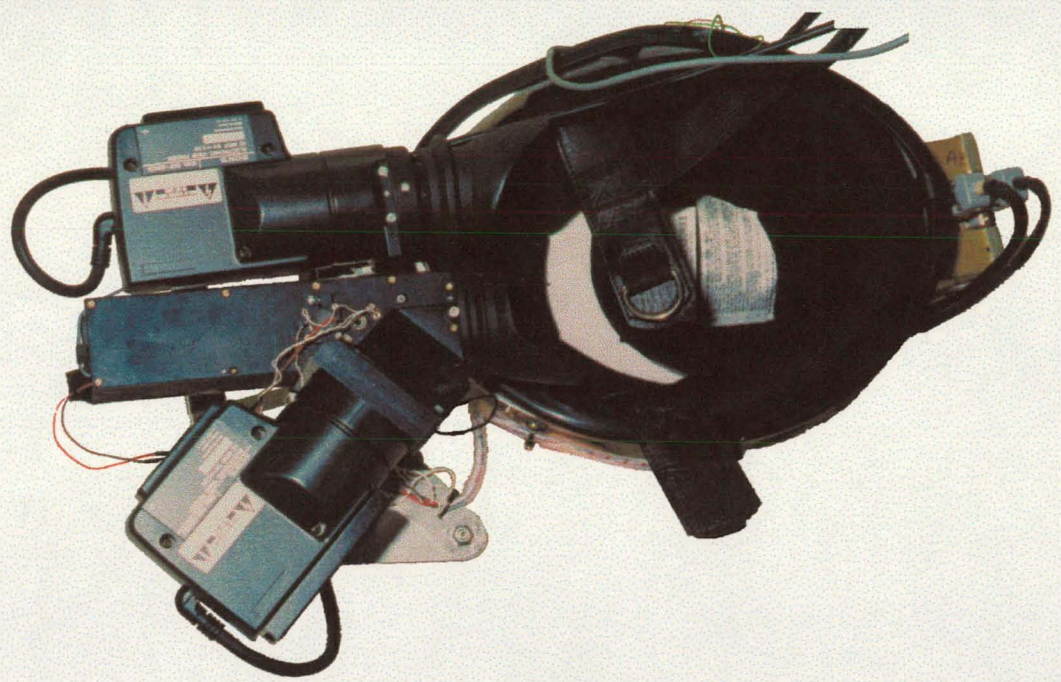


FIGURE 1.9 ESFID DISPLAY

1.5. CHAPTER SUMMARY

This chapter has introduced the research topic, established the thesis focus, and given the physiological and optical motivation for the eye slaved foveal inset display system. The next chapter reviews the applicable literature on displays for teleoperators and target acquisition modelling.

CHAPTER 2 LITERATURE SURVEY

2.1. BACKGROUND

Study and research in the field of remote vision and remote control is a multidisciplinary task since one is dealing with a combination of technology and the human who will observe and react to information provided to him. The disciplines involved range from electro-optic engineering, through visual physiology to perceptual psychology and human factors.

The literature surveyed is in three categories, eye-brain physiology, unaided eye target acquisition performance studies, and remote vision systems. Remote vision is a subdivision of remote control. In the literature, remote control is classified as "teleoperation". Sheridan [6], [36] summarised the current terminology in the field.

a) Teleoperation

The term teleoperator was first coined by E.G. Johnsen and Corliss[37] in 1970. They defined a teleoperator as a general purpose, dextrous, cybernetic machine. The problem with Johnsen's early definition is that it only allows remote manipulation. The concept and the definition of teleoperators were further developed during the past two decades. The most recent formal definition as developed by Sheridan [6] is the following:

> *"A teleoperator is a machine that operates on its environment and is controlled by a human at a distance. Teleoperation refers to the process which is executed by the teleoperator."*

b) Telepresence

Telepresence is the psychological experience when a human operator receives sufficient information about the teleoperator and the task environment, displayed in a sufficiently natural way, so that the operator feels physically present at the remote site.

c) Virtual presence, virtual environment, virtual reality, artificial reality

These terms, which are all synonyms, describe the psychological experience of a person whose sensory information is generated only by and within a computer and associated display technology, and which imparts a feeling of being present in an environment other than that the person actually is in.

2.2. REVIEW OF APPLICABLE LITERATURE

A summary of the available literature and the contributions of other researchers in the specific areas of teleoperator display systems and target acquisition modelling are presented in the following sections.

2.2.1. Teleoperator Display Systems

Johnsen (1972) [37], Vertut and Coiffet (1984) [38,39] and Sheridan (1989) [36] documented the development chronology of modern teleoperators as follows:

- The need for remote vision started in the nuclear industry where teleoperators were used to handle radioactive materials. Direct vision through lead-glass windows and periscopes was used by the operators to observe the remote sites for seven generations of systems from 1947 to 1958.
- In 1958 William E. Bradley, Steven Moulton and associates at Philco Corporation developed a head-mounted miniature TV set that enabled an operator to project himself into the operating space. Comeau and Bryan described this work in 1961[40], and it represented the first experimentation with telepresence.
- 1965. Ray Goertz and his associates at ANL combined the ANL E4 electrical master-slave manipulator with a head-controlled TV camera and receiver. This TV1 system incorporated 2 degrees of freedom, pan and tilt. The success of TV1 encouraged ANL to build TV2 in 1966, possessing five degrees of freedom, including translational motion.
- 1966. The U.S. Navy's CURV (Cable controlled underwater vehicle) was successfully used to retrieve a nuclear bomb from the deep ocean bottom off Polomares, Spain.
- 1967. The lunar teleoperator "Surveyor" collected soil samples from the moon, remotely controlled from earth. It utilised a remote controlled television system.
- In 1968, J. Chatten at Control Data developed a head-controlled television with split foveal-peripheral visual display for moving vehicles [37]. This is the only reference that could be found to split foveal-peripheral displays, but no results seemed to have been published.

The status of teleoperator sensor systems in 1972 was described by Johnsen [37] as follows:

a) Direct viewing systems through special windows for nuclear "hot-cells" and undersea teleoperation.

b) Viewing with mirrors and fibre-scopes into inaccessible places such as pipes and the insides of the human

body.

c) The state of the art remote television-viewing device was ANL's TV2 head-slaved television system. A full size television monitor was slaved by a robotic device to keep it in front of the operator's head. Head position and orientation were measured mechanically. Remote controlled television cameras were also used extensively in space and underwater operations at the time.

d) The first virtual environment display is accredited to Sutherland in 1965 [41]

During the 70's interest shifted towards head mounted displays. At first attention was focused on head tracking devices. Chouet and Young from MIT [42] described an electro-optic head tracking system in 1974 which was intended to give astronauts the capability to control spacecraft attitude with head position. Military interest in head tracking also grew with the intention of giving pilots the ability to aim weapons off- boresight. An example is the F4 Honeywell helmet mounted sight that was used by Shirachi et al. in 1978 [43] to determine head rotational spectral characteristics. These results had a direct bearing on telesensor control requirements as it determined the bandwidth requirements of the attitude control loops of the head controller, which is part of the remote vision system.

Thomas Furness of the Human Engineering Division of the Armstrong Aerospace Research Laboratory at Wright-Patterson Air Force Base in Ohio was one of the pioneers in visual display technology, through the US Air Force's Super Cockpit program. In 1977 he began experiments with the Visually Coupled Airborne Systems Simulator [44]. This "Darth Vader" looking helmet display contained a high definition display of 1200 line resolution achieved through fibre-optic transmission from supercomputer displays. Although the results of these experiments were classified, it is known that the use of eye tracking was incorporated as a means of control. The aim of this research was to generate a virtual cockpit in which the pilot had access to all the information he required in an intuitive and easy-to-use manner.

In the middle 80's Paul Wetzel, M. Eisenman, Ronald Kruk and Brian Welch at CAE developed the fiberoptic head mounted display (FOHMD) with eye-slaved high-resolution graphic inserts [45]. This was used for aircraft simulation. The helmet was extremely heavy and it represented the first experiments with eye-slaved displays.

Just as Chouet and Young proposed to use head movements to control systems, Plesent Goode from NASA

Langley [46] used measurements of the eye's angle to command a pan and tilt television camera. In a teleoperator system this capability allows the natural expansion of the work scene without burdening the operator with trivial cerebral and motor tasks. In 1984 Tachi et al. from Japan's MITI [1] proposed a general anthropomorphic tele-existence system where head and eye measurements are used to present the operator with sensory information from the slave robot as naturally as possible. They built and evaluated a stereo display system with converging cameras and displays with one degree of freedom head movement (pan). This work was followed by the development of a mobile tele-existence system in 1988 [47] where a mobile robot was controlled using their tele-existence display. Their design used quite large display screens that were mounted on an azimuth platform. Manoeuvre tasks with the tele-existence display were completed in a significantly shorter time than with a flat panel two-dimensional display.

Dwight Miller from the Human Factors group at Sandia National Laboratories reported in 1987 [26] on vision systems they evaluated for teleoperated land vehicles as part of the US Army's Teleoperated Mobile Anti-armour project. The variables they used were monochrome and colour and they also tested the difference between simulated and actual remote driving conditions. All the cameras were monoscopic, and panel mounted television displays were used. The cameras were slaved in pan to the vehicle's steering wheels. Their performance metric was the detection range of obstacles. They found that the colour system gave up to 20% longer detection ranges. They also found that the differences between simulated and actual driving operations were small allowing the use of simulators to develop vision systems at much lower cost.

In 1988 Won Kim, Lawrence Stark and their collaborators from the telerobotics unit at Berkeley [25] described a simple, low-cost helmet mounted display for telerobotics. Their design used a two-axis magnetic Helmholtz coil arrangement to measure head position and it utilised two Sony camera viewfinders with 1-inch CRT screens mounted in front of the eyes of the operator. They showed that the helmet display gave consistently better performance over a range of remote manipulation tasks. The design of Kim and Stark formed the basis of the experimental design of this thesis.

At the same time the Naval Ocean System Centre (NOSC) developed the teleoperated vehicle (TOV) and the "Green man" remotely operated, anthropomorphic workstation [48]. Both systems provide a stereoscopic display to the operator using head mounted cathode ray tube (CRT) monitors to give the operators a sense of remote

presence, also called telepresence. The sensing cameras are slaved to the operator's head movements. In the case of the TOV the cameras are slaved over a fibre-optic connection and in the case of the "Green man" through a direct hydraulic coupling between an operator harness, which measures the operator's orientation, and the teleoperator workstation (slave), which mirrors the movements of the operator. These devices were without doubt the most advanced teleoperator systems at the time and they proved the concepts of remote presence and the ability of an operator to project his work over a distance. In 1990 McDonnell and his team at NOSC described their design of a head coupled sensor platform for a teleoperated vehicle [49]. This article gave the design specification which a head coupled teleoperator should meet and can be summarised as follows:

- Angular rate >500 °/s
- Angular acceleration >1000 $°/s^2$
- Bandwidth > 0.5 Hz
- Overshoot must be minimal

These specifications were used as an input to the design of the experimental apparatus of this thesis.

In the late 1980's the whole concept of virtual reality was developing rapidly, fuelled by media hype and visionaries such as Jaron Lanier of VPL Research that was the first company to commercialise the technology. The VPL head-mount design was based on work at the NASA Ames research centre called project VIVED [50]. A large number of academics started serious research on virtual reality and the link between teleoperation and virtual reality. Work done at the University of North Carolina Computer Science Department by Prof. Fred Brooks and Henry Fuchs [51] led to the acceptance of virtual reality as a tool serving science. Early virtual reality researcher Myron Krueger [44] of the University of Wisconsin discovered that people would accept a very simple graphic environment as real. This fact led teleoperation researchers Francis Quek and his co-workers at the University of Michigan to develop the concept of Tele-perception in 1988 [52]. Tele-perception is the segmentation of the computer perception process corresponding to autonomous operations performed on a remote platform and operations performed interactively on a local platform. Computer perception can then be regarded to span a continuum of modalities from full human control to full autonomy. The implications of this research are far-reaching as the display resolution required can be reduced with slave platform computation. Combined with machine vision, the whole new field of model based vision has developed from this concept, which promises to reduce communications bandwidths by a very large factor. The use of tele-perception in display design is a topic for further research.

In 1988 Hiroyuki Yamaguchi and his co-workers at ATR Communications Systems Research Laboratories in Japan proposed a large visual display system that employs eye-movement tracking [53]. Their display concept was based on a back screen projection technique in front of the operator. A large peripheral vision image is projected on the screen, together with an optically superimposed foveal vision image that is slaved to the operator eye gaze point. Yamaguchi used a 14° field of view foveal insert in his work. They used a simple opinion scale to measure the effectiveness of the system, and their findings showed that the system was effective in providing more detailed information to the operator. This is the earliest description that could be found of an eye-slaved display, and the work in this thesis expands on this work in two areas, namely the addition of a helmet mounted system, and a quantitative evaluation technique based on target acquisition performance.

In 1989 Robert Finkelstein, editor of Unmanned Vehicles magazine [54] commented on the utility of teleoperation systems with a feeling of presence, and virtual reality systems, where the images are computer generated, in military scenes in the near future. The man machine interface (master system) of a telepresence teleoperator and a virtual reality system is identical. In the teleoperation case, the sensory information is collected by a remote slave robot while in the case of virtual reality, the sensory information is all computer generated.

Robert Stone of the UK Advanced Robotics Research Centre described the Virtual Environment Remote Driving Experiment in 1991 where they used virtual reality display systems to evaluate human factors in virtual reality and telepresence technologies. They used the VPL EyePhone Model 2 as their display device and the VPL DataGlove as their main control device. Experiments were also conducted with early LCD screens from Casio and Panasonic combined with American LEEP optics. Stone's experiment is one of the first examples where commercial Virtual Reality hardware was used to do serious research on teleoperation.

In 1991 Thomas Sheridan of MIT [55] described telepresence and virtual presence as follows: "Given sufficiently high-fidelity display, a mental attitude of willing acceptance, and a modicum of motor participation, the human operator experiences telepresence (sense of being physically present with virtual object(s) at the remote teleoperator site) or virtual presence (sense of being physically present with visual, auditory, or force displays generated by computer)". There is at this time no theory of presence, let alone a theory of telepresence

or virtual presence. Three determinants for presence were proposed:

a) Extent of sensory information concerning a salient variable transmitted to the appropriate sensors of the observer;

b) Control of relation of sensors to environment (e.g., ability of the observer to modify his viewpoint)

c) Ability to modify the environment

Research still needs to be carried out as to what extent these principle determinants of sense of presence will have an influence on teleoperator task performance. There is also no standard scale to measure presence in a parallel to the Cooper-Harper scale of aircraft controllability that is used in aeronautics.

In the absence of a theory of presence, teleoperation researchers have been following the route of anthropomorphically-designed teleoperators as they offer the best means of transmitting man's adaptive problem solving and manipulative skills to remote work sites. Held and Durlach from MIT proposed in 1991 [56,57] that an anthropomorphical display system can be designed that is slaved to the human operator's eyes. Their proposal was tested in the display design that is demonstrated in this thesis.

In early 1992 the US Army conducted a comprehensive "concept of employment" evaluation test [58] on unmanned ground vehicles called surrogate teleoperated vehicles (STV). The mast-mounted sensor suite design included day and night cameras, an eye-safe laser range finder and designator, a directional acoustic system and an infrared-imaging camera. The display system was a panel mounted television system. Their results are summarised as follows:

a) An optical fibre data link could deliver the high quality transmission that is required over rugged terrain.

b) Telepresence was not sufficiently refined to avoid some fairly dramatic incidents. On several occasions neither operator, nor spectators, could see a hole or ditch large enough to overturn the vehicle. It was also not possible from pure video to determine if the vehicle was on level terrain (perpendicular to down) or on a broad hillside, since it all looks horizontal on a TV screen. Several operators were happy to drive the vehicles over hillside slopes that were so steep that the vehicles turned over! In addition to a NOSC TOV type telepresence system, the operator does require a sense of "down", so that he can develop a feel for the stability or instability of the vehicle. The sense of "down" can be projected in many ways, including stabilised TV cameras that are always aligned with gravity, symbology on the display, and may even be induced by tilting the operator's chair to reflect the remote vehicle's attitude.

The STV tests marked the first time that telepresence was identified as a key requirement for teleoperation.

In 1993 Daryl Rasmussen of NASA Ames [59] reported the results of 2D and 3D manipulation experiments with a head-mounted interface. The study showed that the use of stereo vision and the alignment of the kinaesthetic (the operator's limbic reference) and visual reference frame can reduce task times by a factor two or three. It was also found that more dextrous operations could be performed if the natural hand eye co-ordination of the operator could be used. They suggest that the use of stereo displays and the alignment of kinaesthetic and visual reference frames in laparoscopic surgery will enhance operator precision.

Also in 1993, Andrew Liu, Gregory Tharp, Stephen Lai and Lawrence Stark of the telerobotics unit at Berkeley [60] described the requirements of head mounted displays for teleoperator performance. The major parameters they studied were the update rate (not refresh rate) and delay. For a pick and place task it was found that inexperienced operators require an update rate of at least 10 Hz, while experienced operators could maintain their performance with an update rate down to around 2 Hz. They also found that stereo displays are essential for teleoperation manipulation tasks. If delays are introduced in the command signal to the remote site, the operator's performance reduces linearly in RMS error up to a delay of around 800 ms after which it stabilises. The important result is that any delay that is introduced produces an immediate reduction in performance.

The plethora of Virtual reality spin-off devices is amply demonstrated by surveys of position trackers by Meyer et al. [61] and three-dimensional display systems by McKenna et al. [62] in 1992. In the 1970's position tracking was a very difficult problem that was only used by the military. In the position tracking survey the following technologies were identified as being used (Commercially available examples are given in brackets):

a) Mechanical (Fake Space BOOM of Scott Fisher and Mark Bolas)
b) Optical; Fixed transducer (SELSPOT), Pattern recognition (Honeywell LED array), Laser ranging, Video tracking such as the Elite system of Prof. A. Pedotti at the Politecnico di Milano laboratory and used on the MIR space station.
c) Magnetic; AC Emitter (Polhemeus 3Space), DC emitter (Ascension's Flock of Birds)
d) Acoustic; Time of flight (Logitech 3D mouse), Phase-Coherent (Seitz-Pezaris HMD)

The performance requirements of the tracker in terms of accuracy and latency were also discussed. The

consensus of researchers to date seems to be that latency is more important than the accurate registration of head position [63]. Chung speculates that 100 ms latency is acceptable [64]. Held and Durlach [65] conclude that adaptation is reduced with 60 ms delay, impaired with 120 ms delay and breaks down at delays beyond 200 ms. Stephen Ellis and his co-workers at NASA Ames [66, 67 and 68] have recently shown growing evidence about the acceptability of latency in virtual environments or telepresence systems. The required precision of manipulation is one important factor for performance with latencies. Very short latencies of 16-33ms can be detected during object movement. The total latency requirement in the teleoperation experiment described in this thesis, where there is no manipulation requirement, was specified to be less than 100 ms total.

In a similar vein, Michael McKenna and David Zeltser from the MIT Media laboratory described three-dimensional display systems [62]. Their survey placed most of its emphasis on stereoscopic displays with depth perception, which is not important for the concept described in this thesis, but a number of performance parameters and display technologies were described which are applicable. Table 2.1 gives the performance parameters of the human eye, compared to a state of the art head mounted display.

PERFORMANCE PARAMETER	HUMAN VISUAL SYSTEM	HEAD MOUNTED DISPLAY
Spatial-Resolution (Horizontal x Vertical)	6400x4800	1280x1024x2
Angular resolution (arc-min)	0.5'	4.2'
Refresh Rate	60 Hz	60 Hz
Brightness (cd/m^2 or nt)	10^5	100
Colour (Bits/pixel)	N/A	8 bits each RGB
Information rate	4.3 Mbit/s	3.8 Gbit/s
Field of view	180°x 120°	135°x 90°
Viewing Zone/Volume	360°	360°
No. of views	2 instantaneous	2 generated
Depth resolution	0.47 mm (1 arcmin)	4.0 mm (8.5 arcmin)
Autostereoscopic	Yes	No, requires head tracker

Table 2.1 Human Visual vs. HMD performance

An important observation is that McKenna gives the information rate of each human eye as only 4.3 Mbit/s! This is very low compared to the information rate of 2 Gbit/s of a modern high-resolution monitor and the 5-6 MHz bandwidth of today's TV systems. The eyes are "high-resolution" only in the foveal region, which subtends only about 1-2° of the visual field. A single nerve gives an information rate of around 5 bits/s. Although there are around 5 million cones and 100 million rods on the retina, the receptor information is reduced and processed through five layers of nerve cells before leaving the eye through the optic nerve with approximately 800 000 nerve links (Wandell [2] claims 1.5 million). The eye-slaved helmet design of this thesis addresses the high bandwidth problem of current display technologies, by splitting the foveal and peripheral channels.

The most recent description of display technologies is contained in a study of Woodrow Barfield et al. from the University of Washington [4] that made a comparison between human sensory capabilities and technical specifications of virtual environment equipment. They gave a most detailed analysis of the human visual system in terms of field of view, inter-pupillary distance, optics, alignment, stereo vision, resolution (acuity), sensitivity (luminance and adaptation), contrast sensitivity, colour vision, eye movement, head movement and motion perception. No equipment was described in which a display tracked eye motion, although a number of eye-tracking systems were mentioned. Barfield's study [4] confirms the numbers in table 2.1. It also describes the limitations of head-mounted displays to present visual information at a sufficient level of luminance to match the human's visual capabilities. In this thesis a CRT display was used based on Kim and Stark's design [25], that gives luminance levels up to about 70 nt (cd/m^2), which will be shown later (in paragraph 3.4.1) to adequately stimulate the human eye to its maximum contrast threshold capability.

The most recent work (published in December 97), and which has a most direct bearing on the work described in this thesis, is the work of B. Watson, N. Walker and M. Reddy [69] who evaluated the effect of level of detail degradation in head-mounted display peripheries. To test the paradigm for the design of systems that manage the level of detail in virtual environments, a user study was performed to evaluate the effectiveness of high-detail insets used with head-mounted displays. Ten subjects were given a simple search task that required the location and identification of a single target object. All subjects used seven different displays, varying in inset size (9% and 25% of total FOV of 75.3°(H) x 58.4°(V)), and peripheral detail (25%, 9% and 1% of NTSC resolution). The control display had no inset. The performance measures were the accuracy and search time to identify a target object. No eye tracking was employed. The results showed that insetless, high detail displays did not lead

to significantly different search times or accuracies than displays with insets, for the search task they used. The results indicate that peripheral level of detail degradation can be a very useful compromise for virtual reality computation. Their results also suggest that the effect of eye tracking may be of less importance in a helmet mounted display where the high level of detail inset is not extremely small. Watson et al. used a relatively low-resolution head-mounted display (Horizontal resolution of 6.32 mrad/pixel or 21.73 arcmin/pixel) for the "high resolution" inset with field of views of 18.8°x14.6° and 6.77°x5.2° respectively. In this thesis the inset size will be 8.96°(H) x 6.67°(V) but with a much higher resolution of around 0.24 mrad/pixel (0.825 arcmin/pixel), and the inset is slaved to the eye's position.

2.2.2. Target Acquisition Modelling

The modelling of target acquisition has been pursued in earnest since the Second World War especially by military researchers. The results were most often classified and trickled very slowly into the open literature of the scientific community. The human's target acquisition system is incredibly complex and no researcher to date can describe the full process at neural level, even though an enormous amount of research is being conducted on understanding of the human visual system and the visual cortex [70], [2], [3]. Hubel [70], and many other researchers, have done a large number of studies on animal visual pathways. Shah and Levine [71] have developed the most recent neural models in primate cone pathways, but their work is still on the retinal level. The current status of human target acquisition modelling is that the models are empirical and based on generalised measurements of the human visual system. The development of target acquisition models follows in the wake of the development of an understanding of the human visual system.

The literature concentrates on two aspects of target acquisition modelling [72]. The first is the modelling of eye movement during extended search, and it includes the effect of peripheral acuity. The second aspect is the development of empirical target acquisition models, including the attentional aspects of the visual field, which are used by engineers to design target acquisition systems. These two areas of the literature will be explored separately, although there is a large amount of dependency between the two aspects.

a) Eye-movements and peripheral acuity modelling

It is a well-known fact that the eye exhibits saccadic movement, interspersed with fixations during search. Ogle [27] described this phenomenon in 1950. Ford et al. [73] measured and analysed eye movements during free search in 1959 and came to the conclusion that the mean duration of fixation times was 0.28s. The other important measurement that was made was a frequency distribution of the amplitude of saccades, from which it was found that 86% of saccades are smaller than 15 degrees, and fewer than 1% are more than 25 degrees. The frequency distribution of the size of saccades during naturalistic viewing was measured in Lawrence Stark's lab [74] where it was confirmed that most natural occurring human saccades have a magnitude of 15 degrees or less.

In 1961 Stanley Smith [75] reported the results of experiments where it was shown for the first time that peripheral discriminability up to 26° from the fixation point, had a large influence on visual search time. His studies suggest that those individuals with better peripheral discriminating capabilities would be able to find a target quicker. This dissertation expands on Smith's work as a peripheral vision channel is supplied to the operator to enhance his visual search performance.

Noton and Stark [28] contributed greatly to the modelling of eye-movements for search tasks during the 70's. They discovered that when subjects are provided an experimental situation forcing them to view and recognise previously learned patterns serially, the eye movements tend to follow a fixed path from feature to feature, called the scanpath. They suggested that the very fact that the eye exhibits such scanpaths gives an indication of how patterns are remembered and recognised by all subjects. During occurrences of the scanpath in the pattern learning phase, these memory traces are being laid down. When the scanpath appears in the initial eye movements during recognition, the subject is matching the internal representation with the pattern, by reproducing the successive feature memories, a successful completion of the scanpath indicating recognition of the pattern. Under more normal viewing conditions, when recognition requires few or no eye movements, the eye movements are assumed to be replaced by shifts of an internal attention mechanism, processing successive features of the pattern.

Noton and Stark drew the following conclusions in a follow-up article in 1971 [76]:

i. The internal representation or memory of an object is a piecemeal affair: an assemblage of features or, more strictly, of memory traces of features. During recognition the internal representation is matched serially with the object, feature by feature.

ii. The features of an object are the parts of it that yield the most information (such as the angles and curves of line drawings). It is on these features that the eye fixates during search.

iii. The memory traces recording the features are assembled into the complete internal representation by being connected with other memory traces that record the shifts in attention required to pass from feature to feature, either with eye movements or with internal shifts of attention. The attention shifts connect the features in a preferred order, forming a feature ring and resulting in a scan path, which is usually followed when verifying the features during recognition. This occurred in 65% of the subjects tested. The other 35% remained to be explained.

A probabilistic hypothesis was forwarded by Srinivasan et al. in 1975 [77] to model the regions of high fixation density, which Noton and Stark called features, for any stimulus pattern. The hypothesis was that if eye movement patterns of a subject are observed for a sufficiently long time T, the duration for which the line of sight falls within an element area [$\Delta x\ \Delta y$] around any point (x,y) is equal to $Tf(x,y)\Delta x\Delta y$. The function f(x,y), as defined above is the fixation probability density (FPD). This implies that the regions around the maxima of the fixation probability density function are the ones that tend to be viewed for the largest duration. They further hypothesised that the FPD for a single point is Gaussian, and for a collection of dots it becomes the sum of the contributions from each dot. They expanded the hypothesis to lines and closed and open contours. The hypothesis was tested on the well-known Muller-Lyer illusion, where the distance between open arrows look larger than closed arrows for an identical distance between them. The hypothesis gave a good explanation of this illusion, as the FPD peaks at different distances in this case. A number of other optical illusions could also be explained by this hypothesis. No work could be found which expanded on the work of Srinivasan, even though he achieved impressive results. The determination of the fixation probability density function for a photo image is unfortunately extremely complex, as it has to include the effect of each dot, at every fixation position, thus making it impractical as a modelling technique.

Peripheral Vision:

In the perceptual psychology field, researchers were increasing their understanding of the role of peripheral vision. One example is a study by Jon Leachtenauer from Boeing in 1978 [78], in which he determined a large correlation between photointerpretation performance and peripheral field size. The task was to find targets in aerial photographs, a very realistic target search scene. The peripheral field was determined by a modified Brombach perimeter. The subjects were asked to fixate on a centre point and to view a broken ring at a distance of 33 cm. The subject then had to indicate the direction of the gap, which subtended 15 minutes of arc. The gap was orientated in eight random directions. The ring was enlarged until the gap could not be detected in six of the eight directions, the corresponding angle constituted the subject's peripheral field size. It was found that individuals with larger peripheral fields found more targets per unit time. This relationship was not significantly modified by differences in interpretation experience. It was results such as these that motivated the central theme of this thesis, the addition of peripheral information for the search task in teleoperators.

Search Strategies:

Another question is the effect of search strategies on eye movement. Megaw and Richardson studied the effect of target uncertainty and visual scanning strategies in 1979 [79]. Two sets of experiments were designed, the first forced a specific search strategy on the subject, and in the second, the material itself dictated the strategy. In both series of experiments there has been no evidence of a general lengthening of search times and it was deduced that the role of search strategy was thus not very important for target acquisition performance.

In 1981 Lawrence Stark and Stephen Ellis revisited the scanpath theory after a decade of further research in eye movement and cognition [80]. A large number of researchers established the experimental basis of scanpaths. For instance, Parker (1978) used quite separated but detailed multi-component scenes and asked subjects to detect changes in a scene on representation; his subjects showed the highest incidence of scanpath mode saccades. Parker gives a different interpretation of the ability of subjects to use peripheral vision to that given by Noton and Stark. The problems with the original scanpath experiments were identified as the requirement for a richer set of pictures and scenes, the lack of a quantitative methodology to evaluate the existence of a scanpath or to ascertain the degree of relationship of one example of a scanpath to another. Stark and Ellis presented further experiments with a larger set of pictures, including ambiguous figures with alternative perceptual interpretations. They found that the sequence of fixations in a scanpath could be modelled as a Markov process, as proposed by Senders [81], where the probability of

the next fixation is dependent on the current fixation point. Their experimental results confirmed this theory and it was found that the zero order matrix $\mathbf{M_0}$ represents the percentage time spent by the eye in each fixation region during a search task. The first order Markov Matrix $\mathbf{M_1}$ represents the probability of transition from one of the possible fixation states to the other states. The second order Markov matrix $\mathbf{M_2}$ represents the conditional probability of two eye-movements representing two sequential transitions. That is, an entry in a $\mathbf{M_2}$ fourth-order matrix represents the number of fixations occurring in transition from a state to any of the other three states conditioned upon next but one preceding state. The Markov model gives a quantitative method to determine the existence of a scanpath [82]. The use of ambiguous figures such as the Necker-cube, showed a relationship between the scanpath and the cognitive state of the subject when he is viewing the picture, confirming that cognitive models direct scanpath eye movements in active looking [83]. The exact role of the scanpath in pattern recognition, visual memory and perception is still uncertain.

In 1982 Kraiss and Knäeuper from the Research Institute for Human Engineering in Germany used visual lobe area measurements to predict visual search performance [84] Up to this point in time visual search performance models were based on eye-movement considerations only. Kraiss and Knäeuper suggested that the extent and shape of the visual lobe area was one of the main factors influencing the overall search time in a particular task. They defined the visual lobe area as the peripheral area around the central fixation point from which specific information can be extracted in a single glance. This region is affected by such factors as adaptation level of the eye, target characteristics, background, experience, and motivation. They modelled the visual lobe area with a normal distribution, with the probability of detecting the target on the fovea reducing with the distance between the target and the fixation point. Their model assumed that the target acquisition probability P follows a Poisson distribution:

$$P = 1 - e^{-p_{sum}} \tag{2.1}$$

where

$$p_{sum} = \sum_{i=1}^{n} p(\delta_i) \tag{2.2}$$

and n is the number of relevant fixations, and $p(\delta)$ is the target detection probability from a single glance, given by the normal distribution visual lobe model:

$$p(\delta) = p(0)\, e^{-\left(\frac{\delta^2}{2\sigma^2}\right)} k(\phi) \qquad (2.3)$$

with *p(0)* equal to the probability of target detection at a single glance during foveal inspection, σ is the standard deviation of the visual lobe, δ is the fixation distance from the target, and ϕ is the fixation direction with respect to the target. The authors assumed a circular visual lobe, which meant that $k(\phi)$ is a constant. The authors assumed a fixed grid of fixation points, and achieved relative good results for a simple search situation for this model. They showed that systematic search performance led to an improvement in performance only in cases where the detection probability *p(0)* is high for single glances. To use the model in more complex applications requires the search model to dynamically adapt with various parameters such as field size, visual lobe area, and eye movement characteristics. The problem with Kraiss and Knäeuper's model is that it becomes very complex and requires apriori data sets as each subject, and each subset in a scene, has its own visual lobe [84].

Williams from the University of South Dakota determined in 1982 that the functional field of view of subjects varied with the cognitive load [85]. The functional field of view is defined as the area around the fixation point from which information is being briefly stored and read out during the visual task. This definition corresponds closely with that of the visual lobe, as given by Kraiss and Knäeuper. When given a high level of foveal (cognitive) load, the functional field of view was only about 2 degrees in diameter, whereas a low level of foveal load resulted in a functional field of about 4 degrees in diameter. Williams showed that the shrinkage of the functional field appeared to be rather generalised and not a true tunnel-vision effect. It is this variation in visual lobe size that makes the Kraiss and Knäeuper model impractical.

A number of other researchers also studied the phenomenon of visual lobes in the early 80's. In 1984 Alan Courtney from the University of Hong Kong defined a search task that assesses the visual lobe size [86], in effect turning around the previous work, by using a search task as the input to get the lobe size as an output. Two types of search targets were used, a character "O" and a character "V" positioned off-axis in the picture. He followed up this work in 1985 together with Shou [87], where a simple card-sorting task provided a rapid and simple means of estimating relative visual-lobe size. In 1986 A.P. Golding, from the National Institute of Personnel Research in South Africa [88], used Courtney's method to determine the visual lobe sizes for a large number of subjects. The visual lobe measurements showed a

large variation between subjects, and the visual lobe area varied between 6.34 to 68.66degrees^2, with a mean of 27.36 and a standard deviation of 15.77. Golding also found that the horizontal axis of the visual lobe was significantly larger than the vertical axis, indicating that the visual lobe is ovoid. It is interesting that the isopters of constant visual acuity also have ovoid characteristics, as shown in figure 1.5. Golding found consistent variations in size between the V lobe and the O lobe measurement that corresponds to the type of character used in Courtney's lobe measurement, thus supporting Williams' thesis that the visual lobe varies with cognitive load. The size of the V lobe area was on average 10° horizontal and 5.2° vertical, with 3.0° and 1.95° standard deviations respectively. Golding did further search experiments where he confirmed earlier findings that there is a correlation between visual lobe size, and search times. The correlation was found to be stronger in the horizontal plane than the vertical.

In 1992 Hacisalihzade, Stark and Allen published a paper [89] that studied the stochastic modelling of sequences of eye movement fixations. The visual fixations were modelled both as Markov processes, as well as abstractions of character strings. A quantitative method of measuring the similarity of sequences, based on minimum string editing cost was introduced. The string editing presentation was regarded as more compact and amenable to computation, and applications in machine vision were foreseen.

In 1993 Hung et al. described their measurements of random and direct path eye movements during target search [90]. They found that the majority of the subject's search patterns were random in nature. However in 17% of the cases, the eye moved directly to targets at large initial off-axis angles, thus giving direct path movements. The percentage of occurrence of this behaviour was significantly higher than would be expected by chance. It was speculated that the subjects might make more use of the peripheral information in the visual field. In this thesis the peripheral information is provided in the display to assist visual search in the manner observed by Hung.

In January 1996, Baveja et al. [72] published an optimum overlapping lobes model for visual search, based on the concept of visual lobes. Their model is conceptually similar to that of Kraiss and Knäeuper, except that they had a much more sophisticated model for the inter-fixation distances compared to the simple grid of Kraiss and Knäeuper. They did not use a normal distribution for the visual lobe like Kraiss and Knäeuper, but a simple uniform distribution. Their model was aimed at inspectors and the

experimental field they used was a character field. They achieved good results in determining optimum inter-fixation distances for an inspection task. It is not believed that the visual lobe type of model is of much practical use to predict target acquisition probabilities in complex, highly cluttered natural scenes as used in this thesis, precisely because the visual lobe size varies too much.

To summarise, the Markov model of Ellis and Stark that is based on Senders [81] work, seems to be the most appropriate model for this thesis, as it gives a reliable prediction of fixations for a given scene. The visual lobe models do give an indication of the importance of peripheral vision in target acquisition tasks, which supports the idea of including the peripheral channel in the eye-slaved foveal inset display concept.

b) Empirical target acquisition models

One of the earliest researchers who started to measure the human's visual capabilities under controlled conditions was H. Richard Blackwell [91]. During the Second World War, he did extensive trials (450 000 in total) to determine the contrast threshold of the human eye, as a function of target size and luminance levels. He found that the ambient luminance required for detection decreases as target size increases. His work formed the basis of many target acquisition models to follow.

In 1957 Blackwell followed up his work, together with Alfred Kristofferson and Stanley Smith when they looked at visual detection as influenced by target form [92]. They proposed a theory called element contribution theory, which asserts that the detection of a target occurs whenever the amount of excitation at any point in the neural excitation pattern produced by the target exceeds a critical value. Each point within the neural pattern receives excitation from every retinal point stimulated by the target. To each retinal point there is a corresponding neural locus. The amount of excitation contribution to any neural locus from any retinal point is a function of the distance between that retinal point and the retinal point corresponding to the neural locus. This contribution as a function of retinal distance is the element contribution function. The element contribution theory explained why the threshold targets luminance at which detection occurs, decreases as target size increase. The theory was based on a neural model that at the time was not well understood, and large approximations had to be made. For instance, targets had to be symmetrical and relatively homogeneous. Consequently, the theory did not stand up well to

experimental results at the time. The element contribution theory did however form the basis of the concept of the "perceptual field", which was developed later and used by Srinivasan [77].

In 1960 Krendall and Wodinsky [93] used their extensive experimental results in target detection to show that the cumulative probability of detection in visual search for a target, P_d, follows an exponential time dependence and it could be expressed as follows:

$$P_d = 1 - e^{-mt} \tag{2.4}$$

with

$$m = -\ln(1 - P_g)/t_g \tag{2.5}$$

and P_g is the single glimpse probability of detection, t_g is the glimpse time, and t is the search time. Note the similarity of form with Kraiss and Knäeuper's model as described by equation (2.1), which is based on a Poisson distribution.

John Bloomfield from the University of Nottingham also published an exponential fit model of target detection in 1972 [94]. His research on target disks placed between background disks led to the following amendment to Krendall and Wodinsky's exponential function:

$$P_d = 1 - (1 - P_g)^{(t - t_r)/t_g} \tag{2.6}$$

where the response time t_r is a function of the size of the target and background disks. If the target diameter is d_T and the background disk diameters are d_B, the response time fit he achieved was:

$$t_r = 0.184 + \frac{3.41}{\left|\log d_B - \log d_T\right|} \tag{2.7}$$

Bloomfield's finding that the response time is a function of the difference between the target and the background later led to the concept of a clutter factor.

H.H. Bailey at the Rand Corporation published their first empirical target recognition model in 1970 [95]. Bailey modelled target recognition as the product of the probability that the observer looks for a specified glimpse time (Bailey used 1/3 s) in the direction of the target with his foveal vision (P_{search}), the probability that if a target is viewed foveally for one glimpse period it will be detected, (P_g) and thirdly, the probability that if a target is detected, it will be recognised (P_{rec}):

where η is the degradation factor that depends on signal to noise ratio in the display (see equation 2.13).

The empirical model that Bailey developed for search is:

$$P_{search} = 1 - \exp\left[-\left(\frac{700}{G} \right) \left(\frac{a_t}{A_s} \right) t \right] \tag{2.9}$$

Where G is a congestion, or clutter factor, usually between 1 and 10, a_t the target area, A_s the area to be searched, and t is time.

The detection model for one glimpse period also has an exponential form. The probability of foveal detection at the threshold contrast C_t is by definition 50%. Bailey's model included an approximation for other target/background contrasts C and is given by:

$$P_g \cong 0.5\left\{ 1 \pm \sqrt{1 - e^{-4.2(C/C_t - 1)^2}} \right\} \tag{2.10}$$

The recognition model of Bailey is dependent on the number of resolution cells over the minimum target dimension:

$$N_r = \frac{L_{\min}}{\alpha R} \tag{2.11}$$

where L_{min} is the minimum target dimension [m], α is the angular resolution of the sensor [rad] defined as the number of lines on the sensor divided by the field of view, and R is the target range [m]. The recognition probability follows from Bailey:

$$P_{rec} = \begin{cases} 1 - \exp\left[-\left(N_r / 2 - 1 \right)^2 \right] & \text{for} \quad N_r \geq 2 \\ 0 & N_r < 2 \end{cases} \tag{2.12}$$

Finally, Bailey models the signal to noise ratio (SNR) of the system as a degradation factor,

$$\eta = 1 - \exp\left[-\left(SNR - 1 \right) \right] \qquad \text{for} \qquad SNR \geq 1 \tag{2.13}$$

and if the SNR is less than unity, the target acquisition probability becomes zero.

In 1972 he updated his model slightly [21] to include moving targets and described a "perceived signal-to-noise" concept which may be a better model than that given by equation (2.13). The work of Bailey was included in the RCA electro-optics handbook [20] in 1974 and it became the general reference for

$$P_{acq} = P_{search} P_g P_{rec} \eta \tag{2.8}$$

most electro-optic system designers since the 1970's.

In 1975 Seyb published a detection model that is based on a simple curve fit [96]:

$$P_g = 1 - 0.5^{(F_H CR)^3} \tag{2.14}$$

with

$$CR \equiv C / C_T \tag{2.15}$$

where C is the displayed target contrast and C_T is the contrast threshold of the eye, and F_H is the human factor (≥ 0 and ≤ 1). Note the similarity of the contrast ratio with Bailey's detection model in equation (2.10). Seyb's model has some desirable properties. With the human factor = 1, equation (2.14) approximates (2.10) very well for values of CR>0.5, that is, single glimpse probabilities that exceeds 0.1. In addition, for CR values equal to zero, Seyb's model gives a single glimpse probability of exactly zero, while Bailey's model still gives a value of 0.004. Yet with a uniform target and background, a displayed contrast of zero should imply a target that is impossible to detect in principle - equation (2.14) will produce this result. Another advantage of Seyb's model is that it includes a human factor, which Seyb tells us ranges from unity under laboratory conditions, down to values as low as 0.5 in the field. This additional variable allows the model to account for effects that can degrade human performance, such as fear, fatigue, boredom, or image clutter.

Wessely from the Rand Corporation expanded on Bailey's work in 1977 [97]. He added a line of sight probability model, as well as the probability that the target is in the field of view to be searched, for both single and multiple target scenes. He applied the model to tactical scenes, and used it to show the sensitivity of the limiting factors on target acquisition during a tactical aircraft pass over a target.

In 1978, Lynn Olzak from Hughes completed a study on the detection of multiple-vehicle targets in realistic terrain [98]. Although she achieved good experimental results, she could not fit the results well with any of the existing target acquisition models at the time. She identified the following shortcomings in the algorithms to predict realistic target detection probabilities:

- Target effects – The number and location of the targets had a large influence on performance. Context such as roads aided target acquisition performance significantly.

- Scene characteristics – The characteristics of the scene, such as the clutter factor, lack specificity, and require subjective comparison between scenes, rather than quantitative, objective metrics.

Bernard Koopman did extensive research on target recognition during the late 70's. He published a model for search and detection for a number of realistic targets such as submarines and ships [99]. His initial model for single glimpse detection probability was based on the concept of visual perception angle. Later, he published an empirical detection model [100] based on the integrated Gaussian function and the contrast ratio CR. Koopman's empirical one glimpse detection probability is:

$$P_g = \varphi[2.56(CR-1)] \tag{2.16}$$

with

$$\varphi(u) \equiv (2\pi)^{-0.5} \int_{-\infty}^{u} \exp(-t^2/2)dt \tag{2.17}$$

Koopman's detection model is much more complex than Seyb's model, and for a contrast ratio of 1, Koopman's model still gives a non-zero value. Seyb's model does however give a very good correlation with Koopman's model.

In 1988 John Wootton, Gary Waldman, Greg Hobson and K. Leutkemeyer published a normalised clutter measure for digitised images [101]. It is designed to be in accord with human intuitive estimates of clutter, being based on the similarity of the background texture to the target in size, shape and orientation. They used the joint amplitude histogram, or co-occurrence matrix, defined as a matrix Q of size GxG, where G is the number of grey levels in the digitised image. For each direction A over the image, the element $Q_{ij}(s,A)$ of the co-occurrence matrix is defined as the counts of how often a grey level i is separated from grey level j by s pixels along that direction. Normalising the co-occurrence matrix by dividing with the total number of possible steps of size s along the direction A, gives the transition probability matrix, Pij(s,A). The spread of the co-occurrence matrix about its main diagonal indicates the amount of background texture. Wootton defined the spread function as the absolute value of B for a specific direction A and separation s:

$$B(s,A) = \sum_{i=1}^{G}\sum_{j=1}^{G} |i-j| P_{ij}(s,A) \tag{2.18}$$

For a given direction A, we find that B starts from zero for s=0 and as s increases, B increases to some value at s equal to the average texture size. For larger s, B just gives random variations, as once the step size is larger than the average texture size, any step is likely to land on a different texture element with a different grey level. The knee of this curve amounts to the typical texture size along direction A, and allows comparison with target size. As clutter is proportional to the amount of texture that is similar to the target, Wootton defined the clutter value C for a specific direction A as:

$$C_A = \left(\frac{s}{T_A}\right) B(s,A) \tag{2.19}$$

where T_A is the target size in the direction of A, and s corresponds to the knee of the absolute value of the co-occurrence matrix. The total clutter is the mean of equation (2.19) over a number of directions.

Wootton et al.'s clutter measure has a number of desirable properties. The first is that when averaged over the direction, the maximum clutter is characterised by texture elements that have the same size, shape and orientation as the target. Secondly, the clutter measure is zero for a uniform background, and finally, the metric is independent of the image magnification or orientation. Wootton et al.'s clutter measure thus meets Olzak's requirement for a specific, quantitative clutter metric.

In 1992 Gary Waldman, John Wootton and Greg Hobson published an empirical model of visual detection together with search [102]. They took the best and simplest models that have been developed over the last three decades, and combined them into a field model they called VISDET, which can predict target acquisition performance in practice. They based their one-glimpse detection model on Seyb's [96] model, the contrast thresholds on Blackwell's [91] work, the fixation time on the work of Enoch [103], and the average inter-fixation angle on the work of Enoch and Ford [73]. The clutter measure was based on the work of Bloomfield [94] and Wootton et al.[101]. The VISDET model also included models for the atmosphere, background luminance, and the effect of electro-optical sensors. As Waldman et al.'s model [102] was the best model that could be found in the open literature, it formed the basis of the model that was used in this dissertation.

More complex and complete models exist, but they are either classified, have never been published in their entirety, or are proprietary and thus not reproducible. An example is the Georgia Tech Vision Model (GTV). The GTV model is based on computational models of the image processing performed by the primate visual system, together with what is described as "incorporating many of the findings of basic research on vision, attention, and search". GTV's development was supported by funding from the US Army, Air Force and private industry. It resides as part of a more complex model called VISEO, which is an integrated software environment to design camouflage and signature control systems as described by T Doll et al. in 1995[104].

To summarise, the development of target acquisition models has been traced from 1946 to present. The acquisition models evolved from very simple rules of thumb to very complex models that include models of the neural processing of the human visual system. The latest complex models were found not to be useful, as they were proprietary and they were so complex that they could not be integrated with the system under study. The model that best met the research requirements was Waldman's VISDET model, as it was reproducible and relatively simple to use. As eye movement was included in the research of this thesis, Stark's Markov model for eye movements was found to be of interest. Waldman et al.'s model is described fully in Appendix A, and the model was implemented in MATLAB™, the listings given in Appendix B. The expansion of the model to include Stark's Markov model is also described.

The next section describes the trends that could be detected in the literature. It also identifies gaps that were found in the literature, where more research needs to be done. The contribution from the research in this dissertation is also shown.

2.3. TRENDS AND GAPS IN THE LITERATURE

2.3.1. Research Issues

a) Teleoperator Display Technology

In 1988, Lawrence Stark from University of California, Berkeley Campus [105], defined the main issue that confronted teleoperator display researchers as the man-machine interface. This fundamental

challenge is shaped as much by human perceptual, cognitive, psychological, physiological and motor abilities as by engineering considerations. Mismatches between sensory information and the actual work environment impede performance, and may even lead to simulator sickness [106]. The optical properties of the camera and display have complex effects on the operator's ability to control a teleoperator.

In 1990 W. Uttal from NOSC described the issues which confront display technology as follows:

- Simulator sickness
- Psychological insights in aspects such as telepresence.

Display system issues have been summarised by the work of Barfield et al. [5] who evaluated current virtual environment equipment against human sensory capabilities. They concluded that current visual displays couldn't give the resolution or the luminance range that the human eye is capable of, over its field of view. The work of Tachi [1], who identified the requirement for eye-directed teleoperation, is also supported by Watson et al. who are studying level of detail degradation in head mounted display peripheries [69].

b) Target acquisition modelling

The current state of the art target acquisition models such as the Georgia Tech Vision model (GTV model) [104] are mostly proprietary models, sponsored by the military, and no publication could be found giving the actual models. The GTV model is also too complex to be of practical use. The best simple model that could be found in the open literature was the model of Waldman, Wootton and Hobson [102]. This empirical model assumes that the eye's fixations will form a rectangular grid on a picture to be searched. Various researchers have proven that this is not true. Lawrence Stark has developed the concept of scanpaths, which describes the order in which the human operator searches fixation points in a known picture. The fixation points in a picture are a function of features in the picture, and are difficult to predict. As a target will not easily be detected if it is not positioned close to a fixation point, the fixation pattern that is used will have a large influence on the target acquisition probability. This statement is tested in chapter 5.3. In this thesis the Stark's theory of scanpaths is combined with Wootton's empirical search model, to give a more realistic target acquisition model with which an eye-slaved foveal inset display system can be characterised.

2.3.2. Trend Analysis

In display system technologies the trend has best been summarised by Sheridan [36] in that display systems will become more anthropomorphic and will contain more and more information, up to the point where it matches human vision capability.

In target acquisition modelling the trend is to go for more and more complex models that are incorporating neural modelling of the human visual cortex, as the understanding of the human visual cortex grows. All state of the art target acquisition models such as VISDET and the Georgia Tech Vision model include realistic eye movement. The trend is also that the number of research papers on target acquisition is decreasing with time. Most of the research was done from 1965-1985, where Bailley, Koopmans, Seyb and other researchers were especially active. Publications on target acquisition are currently scarce.

2.3.3. Gap Analysis

No helmet display systems could be found that incorporates eye-slaved displays. The only work that could be found to this effect is the work by Yamaguchi et al.[53] and Watson et al.[69]. This thesis test the concepts proposed by these authors with both a helmet and an eye-slaved display.

No references to split-eye foveal/peripheral displays could be found except for the author's own work [34], and that of Watson et al. who studied lower resolution peripheral displays. References to dichoptic viewing, where each eye is presented with a different picture, abound. The thesis situation is not dichoptic as one eye is presented with a high angular resolution, foveal display, and the other with an identical picture, albeit a wide field, low angular resolution display, which is essentially the same picture, just different resolutions. In this thesis, research results supporting the split-eye foveal/peripheral display concept are presented.

No target acquisition model could be found in the open literature that also supports eye fixation realistically. In this thesis a model is developed based on Waldman's model combined with Lawrence Stark's scanpaths to give a realistic acquisition performance for a known scene. No models of eye movement in a completely unknown scene could be found and researchers are still working on this problem. In this thesis pre-recorded target scenes are used where experimentation determines the fixation points and the sequence in which the operators select the points.

2.3.4. Research Literature Contribution

This research work has contributed to the literature in the following areas:

a) Split-eye foveal/peripheral displays were presented by this author at the Small Satellite and Control Conference in Stellenbosch, South Africa in October 1994. [34]

b) Eye-slaved display target acquisition experimental results were presented by this author at the SPIE AeroSense '98 conference in Orlando, U.S.A. in April 1998 [107]

c) A mathematical model of target acquisition for eye-slaved displays has been developed and is presented in this dissertation.

2.4. SUMMARY OF THE LITERATURE

This literature survey has reviewed the available literature in the fields of teleoperator display systems and target acquisition modelling. The gaps have been identified and it was found that research in split-eye foveal/peripheral displays would contribute to the field of head mounted displays. Furthermore, it was found that high resolution, eye-slaved displays have not been studied before, and that this will also contribute to the field of head mounted displays. A mathematical model that could be used to generalise the research in this thesis was also found. It was proposed to use the VISDET model of Waldman et al. and expand it with the Markov eye-movement model of Stark.

CHAPTER 3 METHOD

3.1. OVERVIEW OF RESEARCH METHOD

The hypothesis on which this thesis is based is described in this section. The hypothesis testing was done experimentally. This thesis is therefore experimental in its approach and it uses mathematical models only to generalise the measurements and to parameterise the proposed solution. The hypothesis is based on a number of concepts that also had to be tested.

By testing the following hypothesis, called Hypothesis 1, the thesis statement of section 1.4.1 will be tested.

"By combining head tracking with eye-tracking in an eye-slaved foveal inset display system, similar target acquisition performance to direct human vision is possible"

During tasks such as picture scanning and recognition, the human eye moves in a series of rapid movements (for 100-150 ms), called saccades, inter-spaced with fixations (for 200-500 ms) on the area being searched. This phenomenon has been described by Noton and Stark [28]. Ford and Stark showed that saccadic movements are limited to less than 15 degrees most of the time [73, 74]. During and for a small time after a saccade, the neural activity of the eye is inhibited [35]. This time window can be used by a display system to track the eyeball, as well as to slave the picture to the eyeball, without disrupting the human visual system.

The hypothesis will be proven experimentally by building an eye-slaved foveal inset display system and testing it with a simulated (pre-recorded) background. A number of subjects will be required to do target acquisition tests for random target positions on a constant background. As a control, the test will be repeated with direct vision, using the identical target and background scenes.

Hypothesis 1 makes certain assumptions, which also have to be tested:

a) Merging Hypothesis

Hypothesis 1 assumes that finite pixel wide-field (e.g. 600 pixels/60°) information and finite pixel narrow-field (e.g. 600 pixels/10°) information can be displayed into separate eyes with the human brain

fusing the two pictures into an apparent 6000 pixel display. This merging assumption is also based on the physiological phenomenon that the eye gives reduced acuity with off-boresight angle, as shown in figure1.6. Therefore, a high angular resolution, narrow-field display and a low angular resolution, wide-field display, if combined, will give close to full acuity over a total field of view of 60°.

The experimental proof will be provided by the cardboard concept tester experiment as described by the author [34] together with an electro-optic split foveal-peripheral display experiment that determines the conditions under which the fusion will occur.

b) Existing Teleoperator Target Acquisition Performance

The second assumption is that current technology teleoperated sensors, which consist of a joy-stick controlled television camera and a display, give inferior target acquisition performance on single military targets, compared to direct human vision. This assumption motivated the development of an improved teleoperated sensor and display design, the eye-slaved foveal inset display system.

This assumption can be tested by doing a target acquisition experiment with direct human vision and with an optimal fixed field of view television monitor with joystick pan and tilt control. The difference in performance is then measured.

This section has defined the hypothesis on which this thesis is based, as well as the experiments that is required to test the hypothesis. The experiments are:

a) A split-eye foveal/peripheral fusion test, using a cardboard tester and an electro-optic tester to show visual fusion,

b) A direct vision target acquisition experiment that will form the target acquisition's control experiment,

c) A terminal display test to quantify typical target acquisition performance of current technology telesensors,

d) An eye-slaved foveal inset display test to demonstrate the concept of eye-slaved display, to determine target acquisition performance using such a display, and to test Hypothesis 1.

3.2. TARGET ACQUISITION EXPERIMENTAL DESIGN

In the previous section the requirement was defined for three target acquisition experiments: a direct vision test, a terminal display test and an eye-slaved foveal inset display test. This section describes the experimental design for these tests. The dependent and independent parameters are defined, and the methods used to assure that all the other variables that can influence the results remain constant through all three experiments are described.

3.2.1. Factorial Design of the Target Acquisition Experiments

Over and above the display method, the target size (range) and inherent contrast are of great importance in target acquisition, as demonstrated by Waldman [102]. To ensure that the full target scene was searched, the experiment was so designed that targets could be placed anywhere in the scene (target position was not known a-priori). Target position was thus added as an independent variable. To summarise, the independent variables for the target acquisition experimental design were defined as the display method, the target range, the target position in the scene, and the target/background contrast. The direct vision test was used to calibrate the independent variable values, and is described in the next chapter.

A 3x2x4x4 factorial design was used, consisting of the three display systems, two contrast ratios, four target ranges and four target positions for the experimental design. The scene was broken up in four quadrants, each corresponding to a target position. The subjects were tested independently from the display method. For each display method a new set of subjects was used to ensure that no a-priori target knowledge could influence the target acquisition test results. The contrast was also tested independently, to minimise target position and range knowledge cross coupling by subjects. This design gave a total of sixteen treatment combinations per test subject. At least 10 test subjects were required for each display type and target contrast to get statistical meaningful data, with a confidence level of around 80-90% depending on the variance. This design required a total of at least 60 test subjects.

The experimental design is summarised in table 3.1. There were 96 treatments in total. The order of presentation of targets was represented by the roman numerals and was the same for each subject. The dependent variable or performance measure was defined as whether the target was actually acquired, or not, in the allocated time. Another performance measure was the time it took to acquire a target in the allocated search time

DISPLAY SYSTEM	TARGET POSITION	CONTRAST LEVEL							
		20% Contrast				30% Contrast			
		RANGE				RANGE			
		250m	500m	750m	1000m	250m	500m	750m	1000m
DVT	Q1	vii	viii	ii	xvi	vii	viii	ii	xvi
	Q2	x	xii	xiv	vi	x	xii	xiv	vi
	Q3	xv	i	iv	xiii	xv	i	iv	xiii
	Q4	iii	v	xi	ix	iii	v	xi	ix
TDS	Q1	vii	viii	ii	xvi	vii	viii	ii	xvi
	Q2	x	xii	xiv	vi	x	xii	xiv	vi
	Q3	xv	i	iv	xiii	xv	i	iv	xiii
	Q4	iii	v	xi	ix	iii	v	xi	ix
ESFID	Q1	vii	viii	ii	xvi	vii	viii	ii	xvi
	Q2	x	xii	xiv	vi	x	xii	xiv	vi
	Q3	xv	i	iv	xiii	xv	i	iv	xiii
	Q4	iii	v	xi	ix	iii	v	xi	ix

Table 3.1 The 3x2x4x4 Factorial Experimental Design

All the rest of the variables had to be held as constant as possible between the experiments. Waldman's mathematical model as described in Appendix A defined these variables as the search time, the scene itself, the size of the scene, the scene display resolution, the targets, the contrast ratio of target to background, the clutter and the light levels. The method used to make each of these variables constant between the different experiments will now be discussed in more detail.

3.2.2. Search Times

The search times were fixed at 30s for longer-range targets (500m –1 000m), and 10 s for the 250m range targets. The 30s search period is derived from the literature. Krendall et al. [93] and Enoch [103] used 30 s maximum search times for an extended search of a complex scene. The 30s period is also supported by SA Army tactical doctrine that allows a maximum of 30s observation time before taking cover during combat operations.

The 10 s search time for very close targets was an arbitrary figure, arrived at through experimentation on the direct vision tester to achieve around a 60% mean target acquisition probability. It is realistic to expect that large and easily visible targets would have a much shorter search time. The chosen search times are further validated in chapter 5, using the actual experimental results.

3.2.3. Background Scene

It was decided to use a pre-recorded digitised background to get a constant scene for the three display systems. The requirements for the scene were as follows:

- The scene had to be the same size for all three experiments,
- it had to have the same angular resolution for all three experiments,
- the scene itself had to reflect typical South African terrain with a medium to high clutter value, and
- the scene had to be recorded with the sun behind the camera, to minimise shadows.

The scene that was used was recorded at 10h30 on 24 July 1997 at De Brug close to Bloemfontein. The azimuth-recording arc is shown on the following topographical map of the area. As can be seen, the sun was directly from behind during the recording.

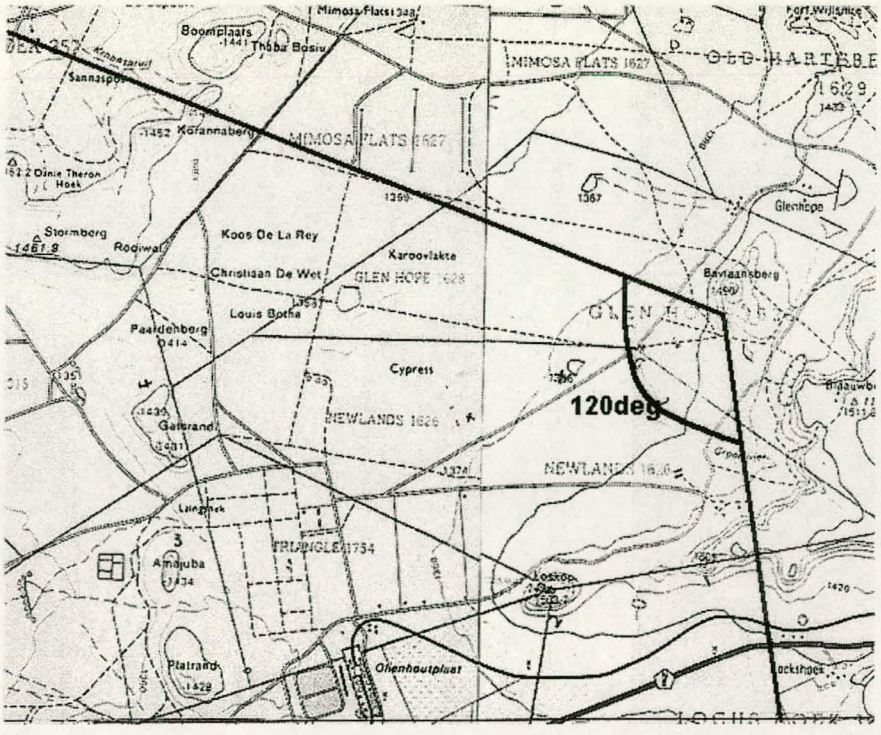


FIGURE 3.1 SCENE MAP

A video camera that was mounted on a computer-controlled pan and tilt platform, which was specially developed for telesensing use, recorded the scene. The platform angles were measured to an accuracy of 1 mrad (3.44 arcmin) using optical shaft encoders. The azimuth angle that was scanned was 120 degrees, and the elevation angle covered 30 degrees. The scene recording was done in steps of 4 degrees in Azimuth and 3 degrees in elevation, to allow a 10% overlap between frames. One frame from each of these steps was then digitised with a Silicon Graphics Onyx computer with a video card. The resolution of each frame was 346 (H) x 263(V) pixels, which corresponds to about half of PAL resolution. The camera was fitted with a zoom lens that was calibrated by a 762 mm measuring rod at a distance of 12.05 m from the camera. The recorded pixel resolution was 0.244 mrad/pixel (0.84 arcmin/pixel).

A computer programme was written to position each of these frames in a scene to form a continuous mosaic. The frames could be integrated with one another to within one pixel over the whole area where targets were placed, which meant that the optical distortion was smaller than a pixel over the recorded field of view. It was also possible to increase and decrease the brightness level of the whole frame to ensure smooth transitions between frames. The final scene that was created was 8192 x 2200 pixels with a 64 level grey-scale. This scene is shown in figure 3.2 at a seven times lower resolution, due to the 600dpi printer limitation. The major landmarks from the map in figure 3.1 are indicated above the scene. The targets were generated using three-dimensional computer graphics, described in more detail in Appendix C and were then placed on the background scene. The scene was split into four quadrants and in each of the quadrants, a 250m, 500m, 750m and 1000m range target was placed. The target positions that were chosen are indicated in figure 3.2, together with their approximate size. This scene, which gave a high clutter value of 0.15-1.5 according to Wootton [101], was used for all the tests. The clutter values were calculated later in the thesis and are given in table 5.2.

The scene used was only a recording, which meant that the absolute value of the target acquisition performance between the actual live scene and the recording would probably be somewhat different due to the loss of colour and contrast from the recording process. The target acquisition experiments were thus done on a relative basis, using the same recorded scene. It will be shown later in the direct vision test that the sensitivity of the absolute target acquisition values were much higher to variables such as target placements than can be expected from the difference between the actual and recorded scenes.

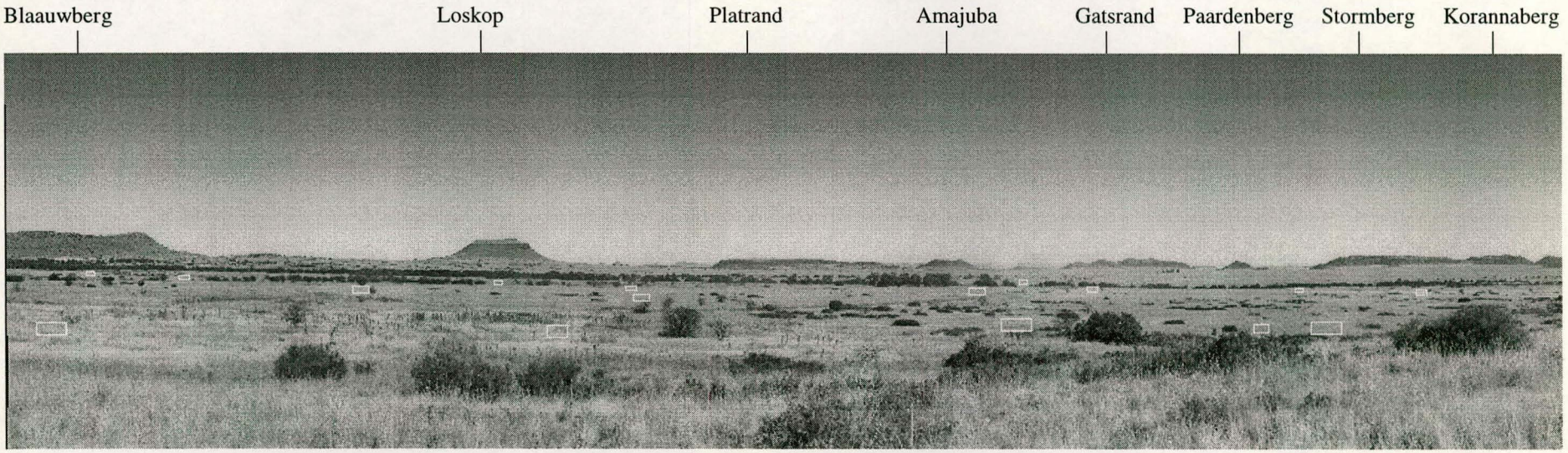


FIGURE 3.2 BACKGROUND SCENE

3.2.4. Target

The target was a 3D model of a T72 tank comprising 322 polygons. The target/background inherent contrast was controlled by the average brightness (colour) of the target. The targets were also generated with the tank facing either to the right or to the left, and the operator was asked to identify the target direction after he acquired the target, to ensure that the correct target was identified. The target was drawn on the scene at its appropriate position and attitude, using a flat shaded lighting model, incorporating both the sun reflection and ambient light conditions. The target was drawn first on a mask, and then on the background after which the whole target was blended into the background with an anti-aliasing filter. The contrast between the target and the background was determined as the average brightness of the target divided by the average brightness of an area surrounding the target with an area equal to that of the target. A typical 30% contrast target at a range of 250 m is shown in figure 3.3.

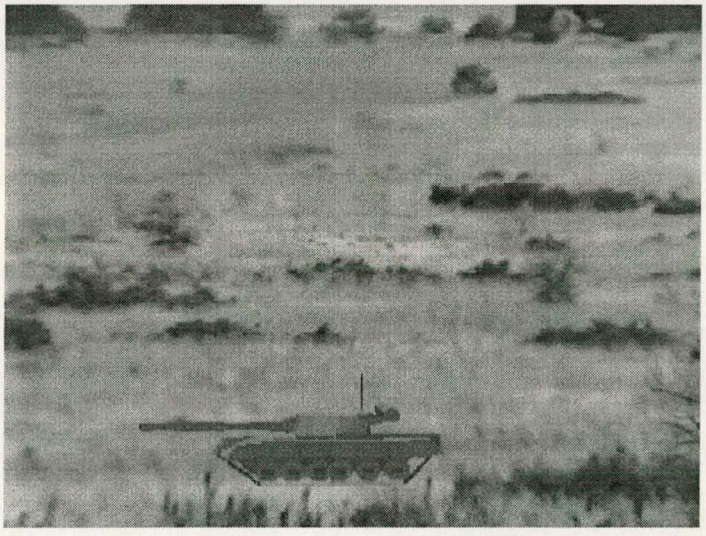


FIGURE 3.3 TARGET MODEL

The contrast threshold depended on the light level according to equation (A.12) in Appendix A [102]. The relationship is shown in figure 3.4. The minimum target dimension was the target height of 2.1m that spanned an angle of 28.9 minutes at 250m range and 7.2 minutes at 1000m range. From figure 3.4 it followed that the contrast thresholds for the targets in our experiments were independent of the luminance level in the 30 to 3000 nt band. The luminance levels for all the display experiments had to fall between 30 and 3000 nt, to ensure that the operators used photopic vision of a similar acuity. The light levels for each of the experiments was measured at the eye position to check that they fell in the required band.

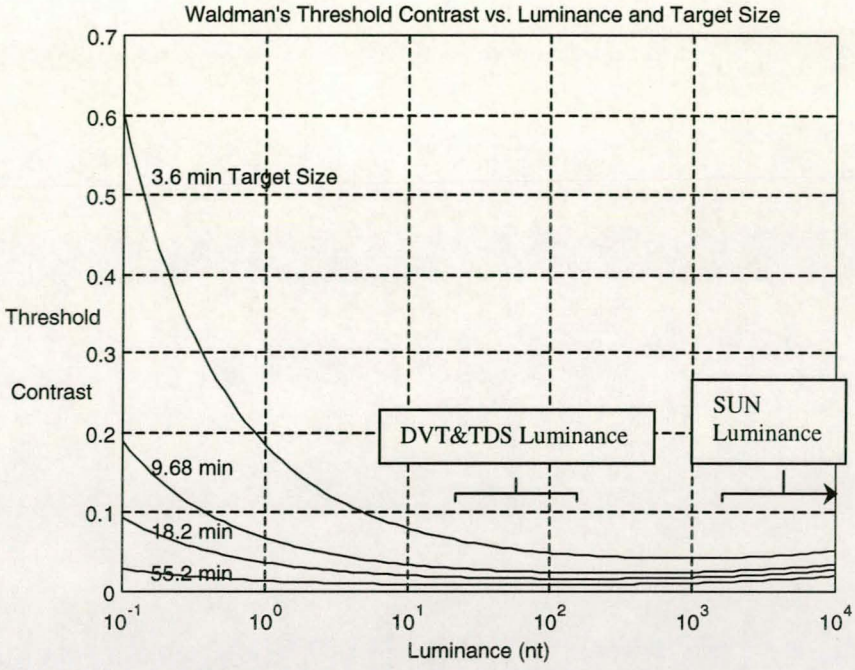


FIGURE 3.4 CONTRAST THRESHOLD VS. LUMINANCE

3.2.5. Test Subject Requirements

The subjects that were used to conduct the experiments were volunteers drawn randomly from an engineering company population. The only requirement posted was that their ages should be between 20 and 45 years. Older eyes often have damaged corneas that can cause a problem for eye tracking. Each subject was tested on a Bausch & Lomb Occupational Vision Tester that provided the following ophthalmologic screening:

- Visual acuity for the left, the right and both eyes, for near and far vision.
- Eye-dominance, as measured by the level of phoria, for both near and far vision. Phoria is the ability of the subject to move the eyes independently [2].
- Peripheral vision, up to 120 degrees.
- Depth (stereopsis) and colour vision were also tested for, although the passing grade was kept low, as these factors are of lesser importance for the experimental design.

Only healthy subjects with 20/20 vision or better (with corrective optics) and no monocular fixation were allowed. As described earlier, monocular fixation is a physiological phenomenon where the subject is suppressing the foveal vision of one eye, usually as a result of childhood strabismus.

The subjects were allowed to train with the apparatus until they had reached a consistent level of proficiency, before the actual acquisition experiment started. The training is described in more detail for each of the tests.

CHAPTER 4 EXPERIMENTS AND RESULTS

4.1. FOVEAL/PERIPHERAL DISPLAY EXPERIMENTS

The purpose of this chapter is to describe the experiments that were defined in the previous chapter, and to present the results of the tests. The purpose of the first experiment is to test the concept of split-eye fovial/peripheral fusion. The following experiments were considered to be necessary to test split-eye foveal peripheral hypothesis:

4.1.1. Direct Vision Fusion Experiment Description

a) Experiment design

The foveal/peripheral fusion phenomenon was studied with direct vision. Ten subjects were selected randomly, and they were not used again in any of the other experiments.

The apparatus is described in the next paragraph (see figure 4.1), and it was designed to allow one eye to receive foveal vision only and the other eye to receive peripheral information only. A relatively long-range target (3-4 km) was selected. The target was big enough for recognition using foveal vision. The subject was placed behind the apparatus and asked if he could fuse the foveal and peripheral pictures into one homogeneous picture.

b) Experimental apparatus description

The first fusion test used the cardboard tester of figure 4.1. The cardboard tester had a cutout for the foveal and the peripheral vision of each eye. The foveal cut-out was 8 degrees in diameter, and the peripheral cut-out 60 degrees. A screen was placed between the cut-outs, to control the viewing distance, and to ensure that each eye received only foveal, or peripheral information. A foveal vision block, of the same diameter as the foveal cut-out, was placed in front of the peripheral eye at the correct interpupillary distance, so as to ensure that only peripheral information was projected into that eye. The cardboard tester was taped against a window pane to allow the foveal block to be positioned directly in front of the eye.

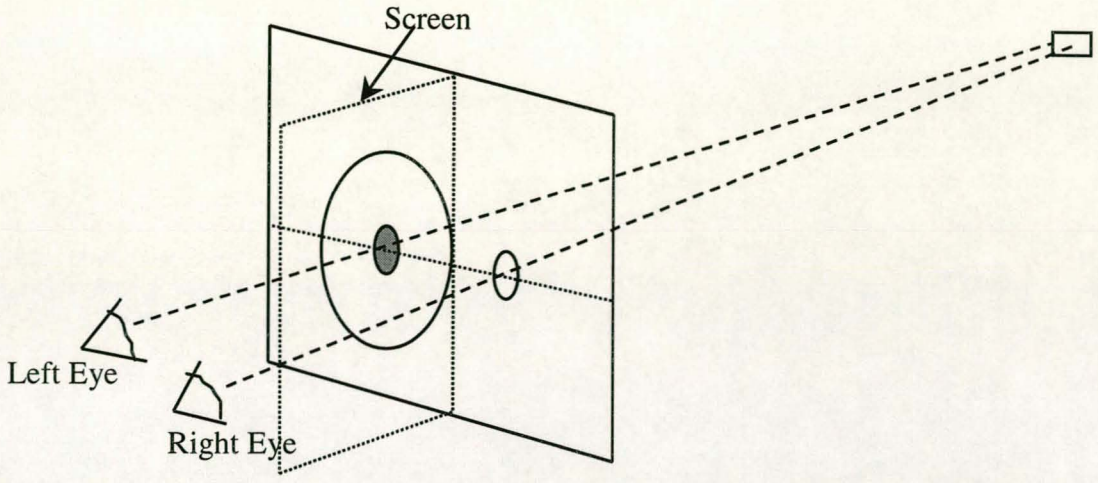


FIGURE 4.1 DIRECT VISION SPLIT FOVEAL/PERIPHERAL FUSION TESTER

4.1.2. Direct Vision Fusion Experiment Results

The direct vision fusion test was done with 10 subjects. They were asked whether they could fuse the wide- and narrow-field displays and make one total picture in their minds. Of the 10 subjects, 8 had no difficulty to do so, while for two subjects it required a bit of training. All the subjects described the foveal picture as much brighter than the peripheral picture. They also described a fuzzy dark ring around the foveal insert. In spite of this, they could all eventually fuse the two pictures into one.

The brightness difference stems from a peculiar physiological fact. The sizes of the pupils are normally identical [3]. With a mismatched light input, as is the case in the direct vision fusion tester, the human visual system averages the pupil size between the two eyes. The mismatch causes the foveal picture to appear much brighter than the peripheral picture. The fuzzy dark ring around the foveal insert can be explained as the blurred edges of the foveal hole, as the cut-out is too close to the operator's eye (100 mm) to be in focus.

The direct vision fusion test was repeated with another 10 subjects. One of these subjects found that he could not keep the display fused for more than a few seconds. All the others experienced no problems in fusing the pictures. These results were presented to Prof. Polla Roux from the ophthalmic department of the University of Pretoria. He was not surprised by the ability of the majority of the subjects to fuse the pictures, and described it as a natural phenomenon. Often one eye loses accommodation before the other. In these cases, the patient still

uses the peripheral vision of that eye and the foveal information of the other eye, to form a highly detailed picture in his visual cortex. The human visual system seems to maximise its information.

One subject could fuse the picture for a short time only, after which the high-resolution picture in his right eye disappeared. A simple eye-dominance test was done and it was found that he was left eye dominant. Eye dominance was determined by asking the subject to point his/her forefinger at an object. The subject is then told to hold his position and then close one eye and then the other. The eye which aligns closest to the object, is the dominant eye. The eye-dominance test was repeated a number of times on different objects. Some subjects used both eyes and others used one eye more often than the other eye - the average is the degree of dominance of the eye. The puzzling fact was that one of the other subjects, who had no difficulty in fusing the wide and narrow field display, also tested left eye dominant. Subsequent discussions with Prof. Roux led us to believe that the phenomenon of monoscopic fixation can explain the discrepancy. Monoscopic fixation often occurs in patients who suffered from childhood strabismus (squint-eye). Their visual pathways were developed with only one eye dominant, and it is not possible for the other eye to initiate fixations. The eye that dominates fixations, is also the dominant eye, and the other eye just follows.

The results of the split-eye foveal/peripheral test led us to the following conclusions about the feasibility of an eye-slaved foveal inset display system.

a) It was found that 95% of the subjects tested could fuse the foveal display into the peripheral display. The one subject that was unsuccessful could be explained as a result of an eye defect called monoscopic fixation.

b) Test subjects have to be screened, both for acuity and eye dominance.

c) The split-eye foveal/peripheral display concept will not work for the following persons:

- Persons with only one eye (trivial),
- Monoscopic fixators, if the high resolution display is not in front of the dominant eye, and
- Persons suffering from strabismus

4.2. ELECTRO-OPTIC FUSION EXPERIMENT

4.2.1. Electro-optic Fusion Experiment Description

The direct vision fusion experiments had shown that observers could fuse a high-resolution sub-image into a wider field of view when both were in perfect angular alignment. The next step was to prove that this could be done with electro-optic displays, and to give the tolerances in alignment and magnification that had to be met for fusion to occur. The electro-optic fusion experiment described in this section was performed to establish these limits.

a) Experiment design

The foveal/peripheral fusion phenomenon was studied with electro-optical displays to determine if the subjects could fuse the two electro-optical displays. The azimuth and elevation angular miss-match, and the magnitude miss-match were the independent variables. A total of 8 subjects were selected randomly.

The fusion was tested in a perfectly aligned electro-optical apparatus. If affirmative, the azimuth alignment between the two displays was adjusted slightly, and the operator asked again to confirm fusion. This process was repeated until fusion broke down, at which point the adjustment was reversed, until the point where fusion again broke down. The positive and negative limits where fusion was lost for each operator were recorded. This process was repeated for the elevation alignment and the magnitude alignment.

b) Experimental apparatus description

The electro-optical display system is illustrated in figure 4.2. A wide field of view display was positioned in front of the left eye of the operator and a narrow-field display in front of the right eye. The interpupillary distance and the azimuth and elevation angle of the displays were adjustable. The wide field display was adjustable in elevation and the narrow field display was adjustable in azimuth. The two displays were provided with video signals from two cameras, a wide field camera and a narrow field camera. The narrow field camera was equipped with an adjustable zoom lens, which was adjusted until a perfect match between the two pictures was found.

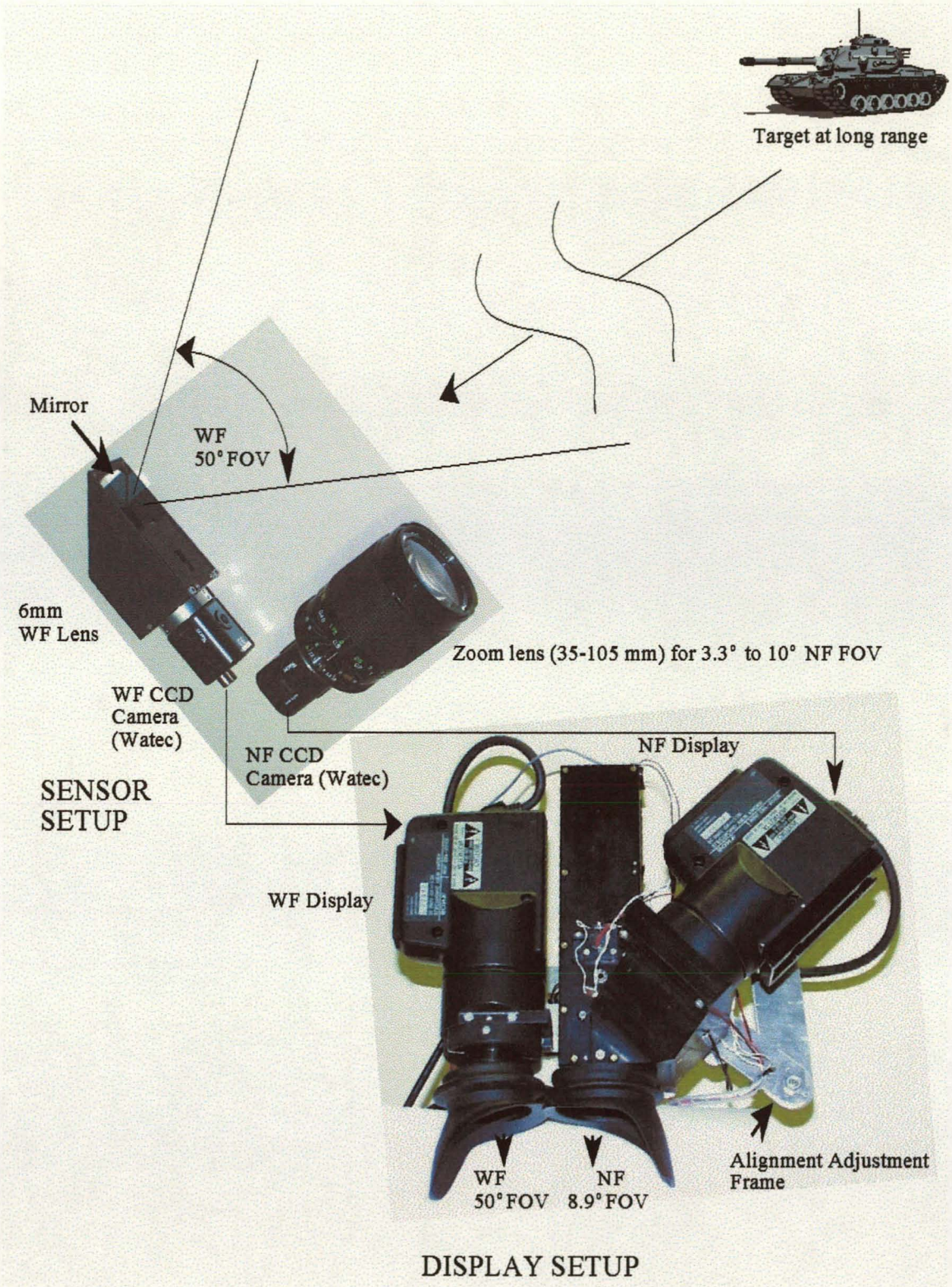


FIGURE 4.2 ELECTRO-OPTIC FOVEAL/PERIPHERAL FUSION TESTER

4.2.2. Electro-optic Fusion Experiment Results

The results of these experiments were reported by the author [34]. Further testing is reported here.

The purpose of the electro-optic fusion experiment was to check if fusion occurs with an electro-optic implementation as occurred with direct vision. It also determined the alignment accuracy necessary for fusion to occur. Eight subjects were tested with the electro-optic fusion tester. All eight subjects had no difficulty to fuse the pictures when the images were aligned. The pictures were then miss-aligned in azimuth, in elevation as well as in magnification, until the subjects could not fuse the pictures anymore.

The results are summarised in table 4.1.

Subject	**Azimuth (mrad)**	**Elevation (mrad)**	**Magnification (%)**
1	+60,-50	+7.5,-7.5	+10,-15
2	+50,-40	+12.5,-10.0	+10,-20
3	+40,-37	+10.0,-12.5	+10,-30
4	+30,-25	+15.0,-7.5	+20,-25
5	+30,-50	+7.5,-7.5	+10,-20
6	+30,-15	+17.5,-20.0	+10,-20
7	+30,-40	+10.0,-10.0	+10,-10
8	+30,-20	+7.5,-7.5	+10,-25
Mean	37.5,-34.6	10.9,-10.3	11.3,-20.6
Standard Deviation	10.9,-12.4	3.5,-4.1	3.3,-8.0

Table 4.1 Split foveal/Peripheral Display Miss-alignment Tolerances for Fusion to Occur

The miss-alignment and magnification data gave the tolerances for the development of the eye-slaved foveal inset display system. The azimuth alignment tolerance was set at ±20mrad (1°8.77'), and the elevation tolerance at ±8mrad (27.5 arcmin). The magnification tolerance was set at ±10% of the field of view. These tolerances were found to be practical for helmet mounted display systems.

4.3. DIRECT VISION EXPERIMENT

The purpose of the direct vision experiment is to act as the control experiment for the target acquisition experiments to follow.

4.3.1. Direct Vision Test Description

The direct vision test was done with the experimental apparatus illustrated in figure 4.3.

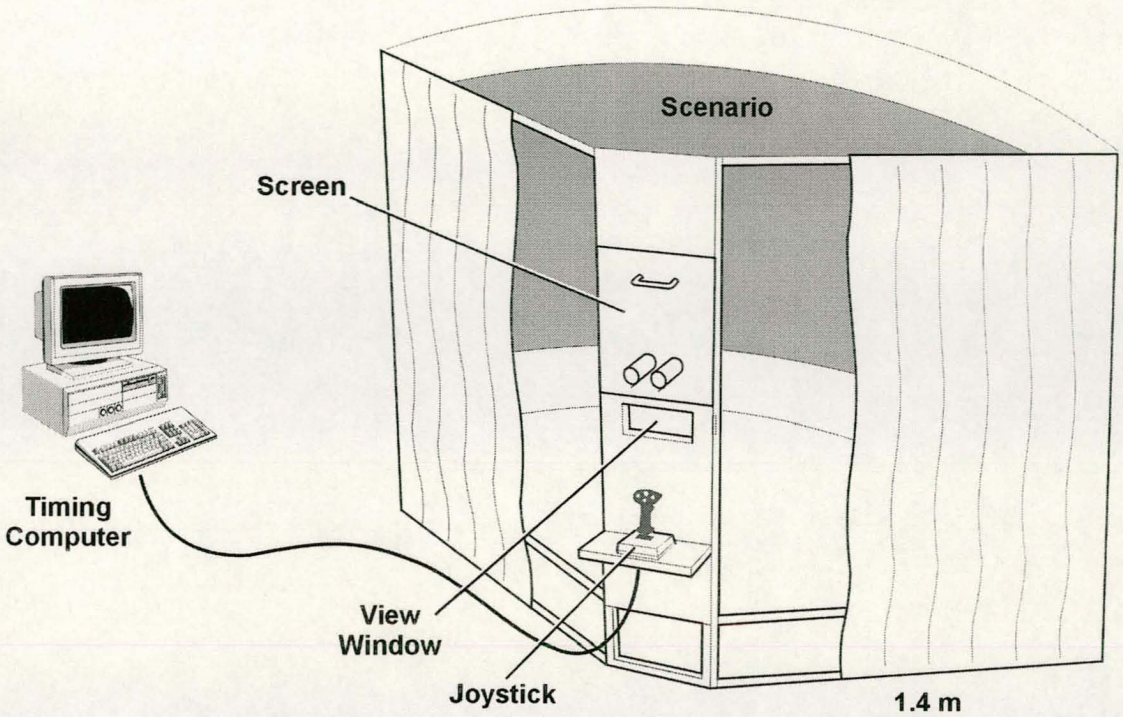


FIGURE 4.3 DIRECT VISION TESTER

The pre-recorded background scene was printed on a 600dpi HP Laser 4 printer. This printer could produce 64 grey levels, which corresponds to the limiting contrast that a human subject can detect in the photopic region [91]. The printer uses an 8x8 pixel size to present the 64 grey levels, which gives a pitch of 75 pixels per inch, or 0.339 mm per pixel. The background scene was recorded at an angular resolution of 0.244 mrad (0.84 arcmin) and to view the printed pixels at the same resolution, a viewing distance of 0.000339/arctan(0.000244) = 1.39 m was required. The scene required a display search area of 120 degrees horizontal by 30 degrees vertical. As the target positions varied little in the elevation plane, a simple wedge shaped enclosure was built consisting of an arced wall with a radius of 1.4 m, flanked by two straight walls. On the inner surface of the arced wall the printout of the scene was pasted as a blended mosaic. The horizontal arc ensured that the angular resolution was constant over the full azimuth display. A screen was placed in front of the operator to blank the scene. Two tiny

holes in the screen allowed the operator to focus on the display before the test commenced. The subject was given a countdown before the screen was lifted, activating a timer at the same time. The subject searched for the target, and when a target was acquired, the target's turret direction had to be indicated by a button. The operator also had to call out the quadrant in which the target was found. This arrangement gave an 87,5% chance that the operator had indicated the correct target. The experimental apparatus is illustrated in figure 4.3.

For each new target position, a new target was generated on the background and printed. The print was pasted on top of the current background with an accuracy that ensured that the print blended into the background. Two fluorescent light units with diffusers provided a luminance level of between 50 to 120nt, as measured at various positions on the scene with a luminance meter. The sky/ background brightness contrast ratio was measured as 180 nt/120nt = 1.5.

For practical reasons, the direct vision tests were conducted in groups of six subjects. More subjects were tested on the DVT than on the other tests. Before each run, a new target was placed in position, and all six subjects performed the experiment, before the next target was positioned. Three training targets were provided to train the subjects in the procedure and to familiarise them with the background scene and the target picture.

4.3.2. Direct Vision Experiment Target Acquisition Results

The direct vision test constitutes the control test for the target acquisition experiments, and it was used to calibrate the targets and the target acquisition performance. It was found that the detailed placement of the targets in the scene had a significant impact on the subject's ability to acquire them. The target placement was refined through the following trial and error method: Groups of six subjects were tested at a time, and then the data was analysed. No subject was used more than once, so no previous target information, apart from the training, could influence the results. Targets that were not seen were moved to more conspicuous locations in that quadrant. It was found for instance that the target's proximity to a bush had a large effect on the acquisition probability. This process was repeated until the mean target acquisition probability for the short-range targets had a reasonably high probability (around 70%), and the maximum range targets a relatively low acquisition probability (around 40%). In the process, the best target ranges were determined as 250, 500, 750 and 1000m. These target positions were also used for the terminal display test and the eye-slaved foveal inset display test.

The target positions that were finally used are shown in figure 3.2. Each target was quite distinct, and no obstruction was modelled, similar to the target indicated in figure3.3.

The tests were conducted with groups of six subjects at a time. The 20% contrast target tests actually ended up with thirteen subjects due to scheduling mistake. The 30% contrast targets were tested with twelve subjects. All the subjects had no previous exposure to the direct vision test. Three training targets were used to familiarise the subjects with the target indicator controls and the time constraints. Each person was allowed to use his own search method. The majority used a systematic scan from left to right.

The direct vision target acquisition results are summarised in table 4.2 and 4.3 for the 20% and 30% contrast targets respectively. The first three targets were training targets and their results were ignored. The run number in tables 4.2 and 4.3 corresponds to the target conditions in table 3.1. The results are given in terms of the time required to acquire a target, in seconds. A blank indicates that the target could not be found in the allocated time. False targets are indicated with a "FT" symbol.

The fact that there is no learning curve apparent in the results of table 4.2 and 4.3 show that the subjects have reached a performance asymptote during training. The tests were conducted in the order shown, and there is no strong evidence that more targets were acquired later in the run.

Subject No.	RUN NUMBER																No. of Targets Acquired
	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Quadrant	3	1	4	3	4	2	1	1	4	2	4	2	3	2	3	1	
Range	500	750	250	750	500	1000	250	500	1000	250	750	500	1000	750	250	1000	
32	1.81	18.62	2.31	8.51		23.5	2.2	7.08		4.94		2.31		2.3	1.32		11
69	14.28							17.63		2.69	14.11			7.19	2.08		6
77	7.97						5.11	8.79		2.09	14.28	1.87		5.88			7
94	3.51			24.11		19.5	9.5	8.29				20.16		3.46			7
56	16.91				29.06		2.58	11.59		2.69		7.58		4.94	7.53	28.5	9
76			8.56	27.3		24.2		9.34		4.94		14.94	4.17	11.7	6.81		9
87		20.66			21.47		7.96	12.31		5.27		6.54	29.9	8.4			8
36			5.88				5.38	5.28	20.1			15.05		6.32	5.6		7
100	5.05			6.75		11.92	6.15			4.28	26.5	2.25	13.56	2.09	1.76	7.58	11
42	9.78	13.29	6.26	18.95	4.01	29.77	3.18	6.59		4.41	19.72	10.27		10.27	4.56	10.21	14
110	2.86					27.24				2.48		18.29	26.86	19.0	3.89	14.72	8
126	11.15	20.9	8.9		29.9	6.7	7.1	12.8	29.9	7.1	5.99	2.42		4.78	6.65	11.15	14
124	11.43		9.9		2.14		2.59			6.2		10.44	5.05	11.92	4.56		9
Mean	8.48	18.37	6.97	17.12	13.42	20.4	5.17	9.97	25.0	4.28	16.12	9.34	15.9	7.55	4.48	15.03	9.23
Std. Dev.	4.84	3.06	2.52	8.22	12.0	7.7	2.4	3.48	4.9	1.56	6.79	6.3	10.74	4.55	2.1	7.5	2.48
Pacq	0.77	0.31	0.46	0.38	0.38	0.54	0.77	0.77	0.15	0.85	0.38	0.92	0.38	1.0	0.77	0.38	0.577

Table 4.2 Direct Vision Results for a 20% Contrast Target

Subject No.	RUN NUMBER																No. of Targets Acquired
	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Quadrant	3	1	4	3	4	2	1	1	4	2	4	2	3	2	3	1	
Range	500	750	250	750	500	1000	250	500	1000	250	750	500	1000	750	250	1000	
140	5.28		2.92		11.86		9.51	2.53		6.54		9.39	26.15	14.89	1.48	11.59	11
21		29.9					7.52	3.57	19.88	6.75	15.16			13.46	3.96		8
108		18.29						8.95		9.9		11.75		8.68			5
143	5.44			28.5	25.8	5.11	2.53	5.0		1.97		1.92		24.5		25.29	10
3	1.82		2.14	10.49	5.93	4.23	6.76	2.74		5.93	15.27	3.02	29.9	11.48	3.02		13
145	10.93							5.6		4.83		26.09		6.81		18.18	6
147	29.9					7.91	1.81	3.79		3.13		7.25	18.67	6.64	6.27		9
97	22.08						8.96	8.07		2.31				24.22	6.53		6
130	5.33	10.05		12.53		8.46	2.04	2.58	2.69	4.77		14.12	23.28	8.02	5.54	10.48	13
47	17.14		3.24				1.87	23.94		1.43		17.85		16.97	5.39		8
7	10.93	23.29			6.98		4.28	22.08	10.11	3.41	6.81	23.95		16.43			10
148	6.37	29.9				5.93	4.07	18.68		1.26	29.9	23.62	23.13	7.96		20.1	11
Mean	11.5	22.29	2.77	17.17	12.64	6.33	4.94	8.96	10.89	4.35	16.78	13.9	24.23	13.33	4.6	17.13	9.16
Std. Dev.	8.46	7.5	0.46	8.05	7.92	1.62	2.86	7.6	7.03	2.49	8.31	8.3	3.71	6.04	1.72	5.5	2.54
Pacq	0.83	0.42	0.25	0.25	0.33	0.42	0.83	1.0	0.25	1.0	0.33	0.83	0.42	1.0	0.58	0.42	0.573

Table 4.3 Direct Vision Results for a 30% Contrast Target

4.4. TERMINAL DISPLAY EXPERIMENT

The purpose of the terminal display system experiment was to test the assumption that current technology teleoperated target acquisition systems gave inferior target acquisition performance compared to direct vision.

4.4.1. Terminal Display System Test Description

The terminal display target acquisition experiment used a simulation of a sighting system in which a limited field of view was displayed to the operator on a video monitor. This limited field of view display was steered by the operator, using joystick input. This experiment simulated conventional technology television-based remote sighting systems. The scene displayed was identical to that recorded for the direct vision scene. The difference was that only 640 by 480 pixels were displayed at any one time. The operator was positioned at a distance of 1.83 m from a monitor (0.35m diagonal), which corresponded to an instantaneous field of view of 8.9 degrees horizontal by 6.6 degrees vertical. This ensured that the operator's on-screen visual acuity matched that of the recorded scene being displayed, that is 0.244 mrad/pixel (0.84 arcmin/pixel). The instantaneous field of view angle was chosen on the following basis:

a) The field of view was the same as that of the eye-slaved display, allowing direct performance comparisons to be made. The teleoperator field of view was then a constant between the displays, and was not regarded as an independent parameter in the experimental design.

b) According to Enoch [103], who has done research on the effect of display size on visual search, the optimum size static display would subtend 9 degrees at the eye. Smaller display sizes lead to lower search performance due to the eye's tendency to fixate on regions outside the display. Enoch found that larger display sizes also lead to lower performance due to nonuniformities he found in the distributions of eye fixations on the display.

c) A number of existing sighting systems use field of views close to 8.9° x 6.6° [96]:

- McDonald Douglas MMS mast mounted observation sight (Kiowa Scout helicopter) 8°x8°
- Israeli IAI TAMAM SLOS (Stabilised long range observation system) 9°x9°
- Israeli TADIRAN MOKED electro-optical sensor for UAV's 10.75°x7.25°
- Martin Marietta Pathfinder Pod for F16 Navigation, TV FOV 9°x7°

d) The field of view is consistent with PAL TV viewing distance guidelines of 0.15 to 0.2m per inch screen size (diagonal) [33].

The experimental apparatus consisted of a personal computer, joystick, RGB to TV converter and a video monitor, as illustrated in figure 4.4. The joystick translated the operator's inputs into rate commands to the sight simulation. The output of the sight simulation was the viewing angles of the sight. The pre-recorded scene was loaded into the extended memory of the computer. The sight angles defined which part of the scene had to be copied to display memory. The joystick commanded the line of sight rate. The joystick gain K in the block diagram in figure4.4 was adjustable by the operator to enable him to search the full area inside his allotted time period. The simulated sight position, θ_{LOS}, was determined by passing the joystick signal through a second order filter with a bandwith ω of 10 Hz and damping ξ of 0.7. The appropriate position in the extended memory of the computer was then copied to the display screen. Target acquisition was indicated by pressing a button on the joystick, similar to the DVT. The terminal display system experiment is depicted in figure 4.4.

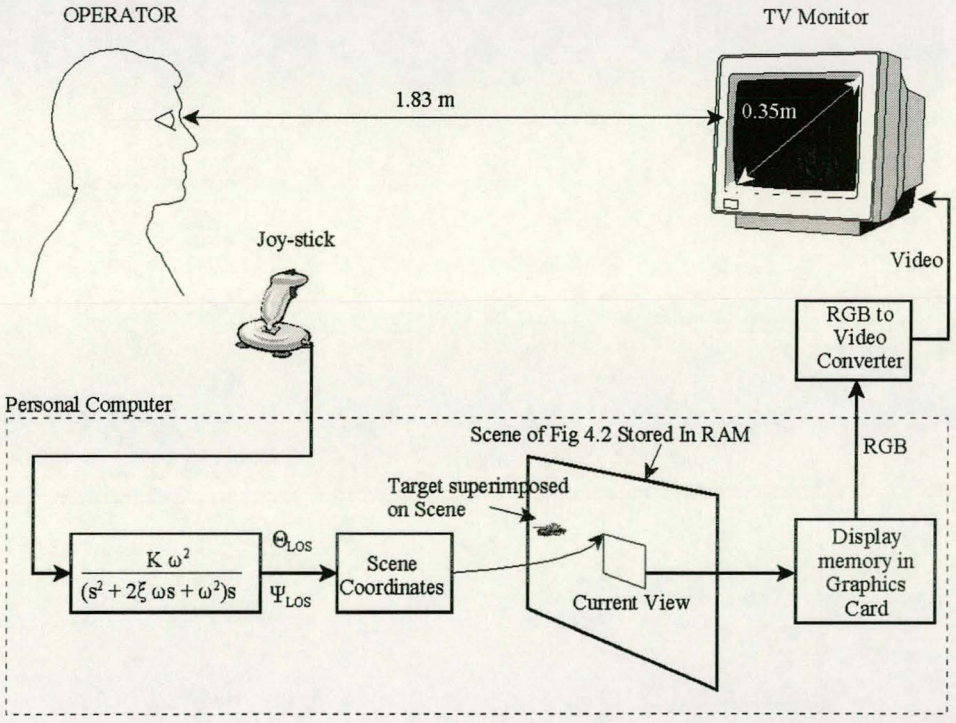


FIGURE 4.4 TERMINAL DISPLAY EXPERIMENT

The target was added electronically to the scene at the start of each session, using the target model described in the previous chapter. The lighting model in the target used field measurements. The brightness and contrast were set until a test pattern of 64 grey scales could be displayed, to ensure that the target contrast between the direct visual display and the terminal display were equivalent. The measured luminance from the monitor on various

positions of the scene on the terminal display varied between 30 and 140 nt, which was within the experimental design's specified range.

4.4.2. Terminal Display System Results

The terminal display experiment was conducted with 12 subjects for the 20% contrast target, and 10 subjects for the 30% contrast. On average, each subject received about 20 minutes training using the three training targets. Most of the training went into familiarising the subject with the joystick controls, and to develop their search strategy within the time constraint. Subjects were allowed to adjust the joystick gain to where they were most comfortable, but with which they could still search the whole scene inside the allotted time. The actual testing against the sixteen targets took another 20 minutes on average. The subjects had no previous exposure to the terminal display tester, or the direct vision experiment.

Each run started exactly in the middle of the scene, with the crosshairs on the horizon. The operators used the search strategy with which they were most comfortable. A number of different strategies were used, including:

a) A quick scan to the left/right limit, and then a constant scan back.

b) A scan to the left/right, followed by a quick scan back to the middle, before scanning the other half

c) A joystick controlled step-scan pattern with overlapping screens, searching only stationary screens.

d) One subject used a zigzag up and down motion, combined with a constant azimuth slew- rate.

The results are summarised in tables 4.4 and 4.5 for the 20 and 30% contrast targets respectively. The subjects reported more difficulty with the terminal display test than the subjects of the direct vision test. The oral feedback was the following:

- The targets were very difficult to see, especially while the sight was panning.
- The control was done mostly in the azimuth plane, and only a few subjects searched in the elevation plane.
- The ten seconds that was allowed for the 250m target search was too short if the joystick was initially sent in the wrong direction.

The experiment was not changed as a result of this feedback, as these problems are inherent to the terminal display concept, which was exactly the objective of the test.

Subject No.	RUN NUMBER																No. of Targets Acquired
	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Quadrant	3	1	4	3	4	2	1	1	4	2	4	2	3	2	3	1	
Range	500	750	250	750	500	1000	250	500	1000	250	750	500	1000	750	250	1000	
49	21.1							10.65	28.6			6.2		6.3		15.0	6
75								8.68						24.39		27.24	3
111							7.41		FT					26.64	6.1		3
121		FT						FT									0
137	29.1			29.39				12.6				8.78					4
138	26.37											18.18					2
58	4.01							25.15								13.87	3
107	24.55							24.77			10.98	21.09		19.23	4.23		6
154	21.64						7.85	13.9				15.76		14.44			5
162	22.91															FT	1
157	22.47							25.26				24.39		25.1	5.93		5
166	20.93	20.92								6.37		4.67		4.56			5
Mean	21.45	20.92	-	29.39	-	-	7.63	17.28	28.6	6.37	10.98	14.15	-	17.23	5.42	18.7	3.58
Std. Dev.	6.67	-	-	-	-	-	0.22	6.9	-	-	-	7.1	-	8.4	0.84	6.05	1.85
Pacq	0.75	0.08	0.0	0.08	0.0	0.0	0.17	0.58	0.08	0.08	0.08	0.58	0.0	0.58	0.25	0.25	0.224

Table 4.4 Terminal Display System Results for a 20% Contrast Target

Subject No.	RUN NUMBER																No. of Targets Acquired
	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Quadrant	3	1	4	3	4	2	1	1	4	2	4	2	3	2	3	1	
Range	500	750	250	750	500	1000	250	500	1000	250	750	500	1000	750	250	1000	
127	25.8		8.02		9.39			24.67				5.1		5.94			6
164	3.07							18.8	21			3.57	3.46	3.68			6
156			FT		FT		FT	FT	FT		FT						0
11													5.3				1
132			8.85											20.33			2
170						6.64											1
93	18.95			21.97					25.43	5.65		10.59		4.34		10.99	7
134																	0
149	2.96					FT	9.67	19.22				3.59		28.4	3.35		6
19										6.55							1
Mean	12.7	-	8.44	21.97	9.39	6.64	9.67	20.9	23.2	6.1	-	5.71	4.38	12.54	3.35	10.99	3
Std. Dev.	10.0	-	0.42	-	-	-	-	2.67	2.2	0.45	-	2.88	0.92	10.0	-	-	2.72
Pacq	0.4	0.0	0.2	0.1	0.1	0.1	0.1	0.3	0.2	0.2	0.0	0.4	0.2	0.5	0.1	0.1	0.188

Table 4.5 Terminal Display System Results for a 30% Contrast Target

4.5. EYE-SLAVED FOVEAL INSET DISPLAY EXPERIMENT

The purpose of the eye slaved foveal inset display experiment is to test Hypothesis 1, that is that by combining head tracking with eye-tracking, similar target acquisition performance to direct human vision is possible. The target acquisition test results for all three testers have been presented at SPIE's Aerosense'98 conference in April 1998 in Orlando, U.S.A. [107], as part of their helmet mounted display conference. Since then further tests have been conducted, which are also reported here.

4.5.1. Eye-slaved foveal inset display system Test Description

The eye-slaved foveal inset display simulator block diagram is shown in figure 4.5. A detailed description of the design can be found in Appendix C. The experimental set-up consists of the eye-slaved display helmet, an eye

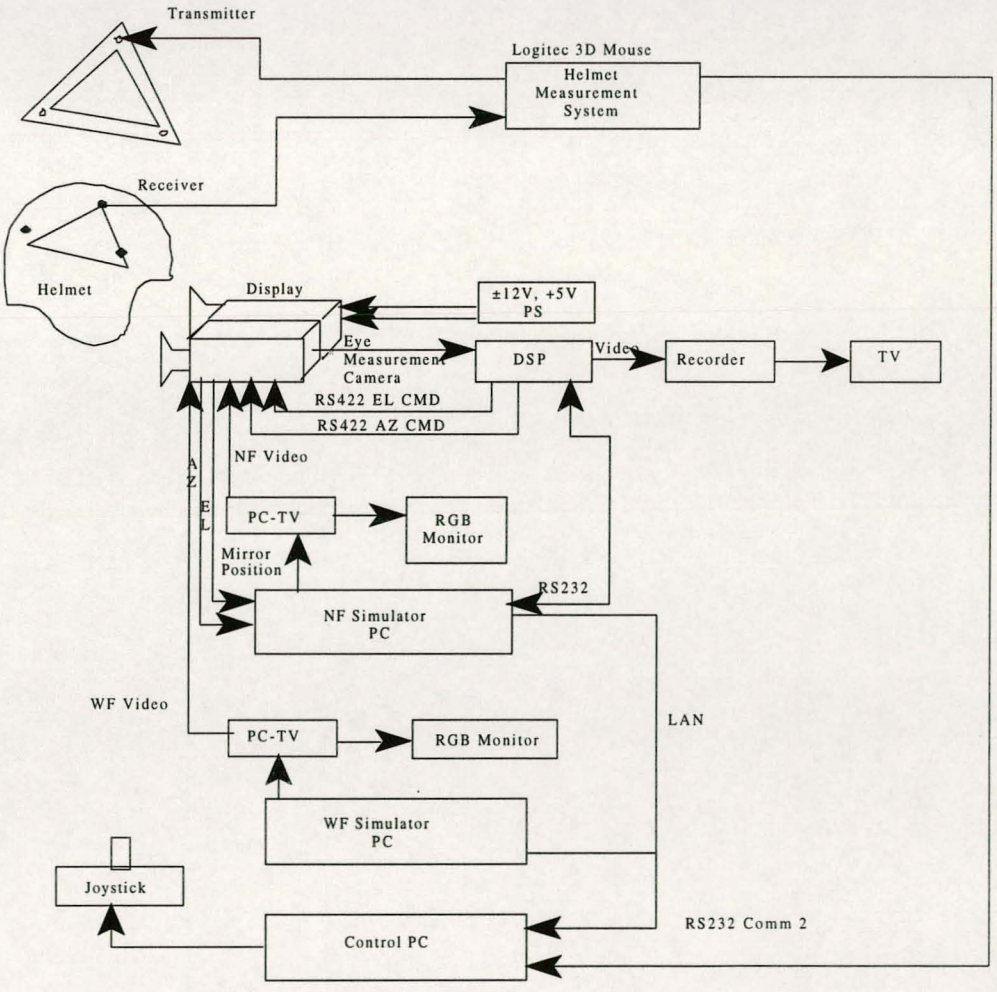


FIGURE 4.5 EYE-SLAVED FOVEAL INSET DISPLAY SYSTEM BLOCK DIAGRAM

tracker, an ultrasonic helmet tracker, a control computer that controls the experiment and two graphics computers to generate simulated pictures in the wide and narrow field channels respectively.

The eye-slaved display helmet design is shown diagrammatically in figure 4.6. The wide field of view display is fixed to the helmet, with a 50°x37.5° field of view. The narrow field display is slaved to the operator's eyes. The eye-slaved control loop comprises a camera with reference 900 nm LED, DSP hardware and software and two servoed mirrors that deflect the display in azimuth and elevation respectively. The narrow field of view subtends 8.96°x6.67°, the same as that of the terminal display. The field of view is achieved with a 150mm focal length objective lens on a 1" CRT, which has a display area of 23.5x17.5mm. The azimuth mirror is dichroic, which reflects visible light and passes near infra-red for the eye tracker reference LED and camera.

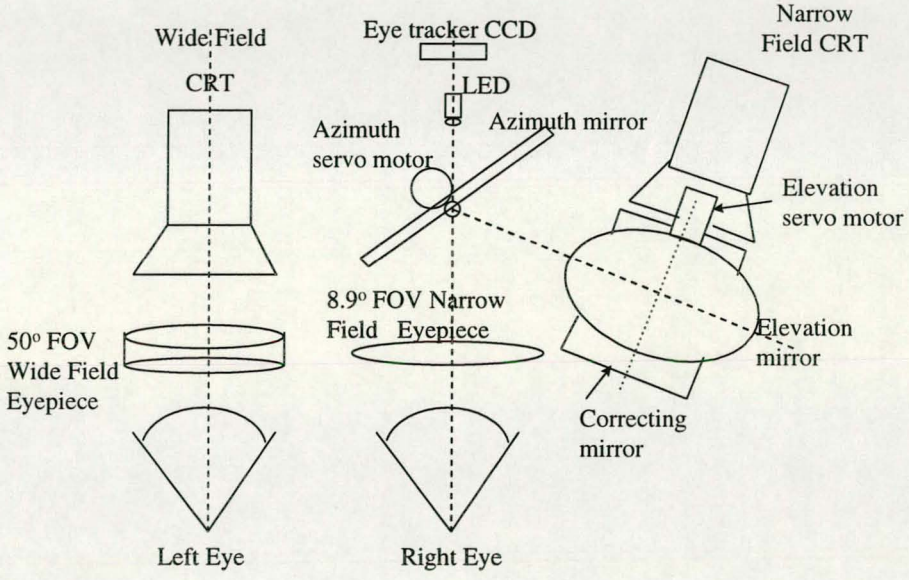


FIGURE 4.6 THE EYE-SLAVED DISPLAY HELMET DESIGN

The control of the eye-slaved display was done by B. Nortje, who worked with the author on the development of the display. The control system is described in detail in Nortje's Masters Thesis [110], and only a high level description will be given here. The eye tracker uses DSP on the CCD video signal to determine the centre of the pupil, as well as the edge of the LED reflection from the cornea. The DSP algorithm processes the typical frame of the eye tracker camera, grabbed by the DSP as shown in figure 4.7. The algorithm first searches the picture for the saturated LED reflection (A in figure 4.7). The LED power can be adjusted to ensure that the reflection is saturated. The algorithm assumes that ten pixels to the left of the LED reflection, there is a pupil, and then it searches for the bottom of the pupil (A1). Once found, the algorithm moves back ten pixels and runs an edge

detector along that line towards the left, that is an edge that goes from dark to light (B). Once the edge is found, the algorithm searches up in the pupil for the edge indicated by C. The average vertical distance between edges B and C gives the centre of the pupil in the vertical. Depending on the position of the LED reflection, the algorithm either runs on the top or the bottom edge towards the right to determine the horizontal edge of the pupil on the right hand side. The search starts 30 pixels towards the left of the LED x position A and searches towards the right for a dark to light edge (D). Even if the search starts on the left side of the pupil, it is still not a problem. The centre of the pupil is found by halving the horizontal line. This algorithm is independent of the pupil diameter, so changing light levels does not influence the eye tracker.

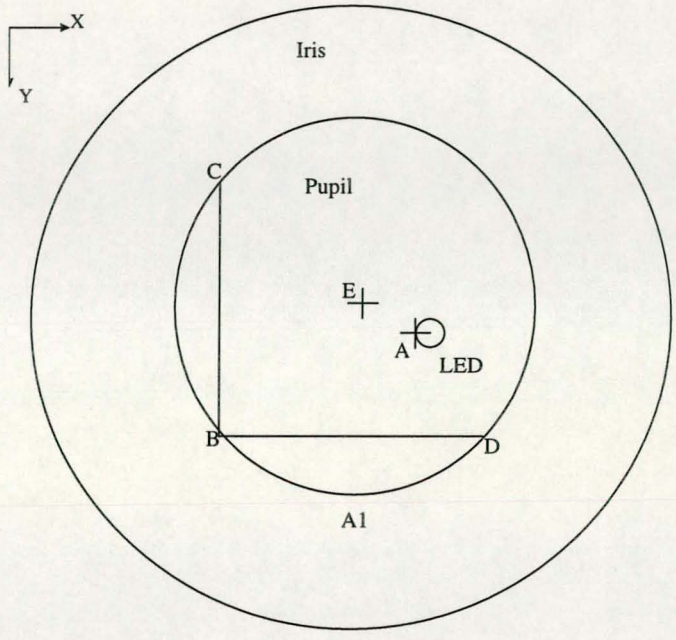


FIGURE 4.7 EYE TRACKER ALGORITHM

The centre of the pupil is given by E in figure 4.7, and the difference between A and E for both the X and Y directions are the eye tracker's raw measurements. These measurements are given to a neural network that determines the look angle of the eye. The units used at the moment are levels, which corresponds to about 200 levels per degree, using a 12bit digitiser. A dead band of 0.5 degrees and a slew rate filter is added at the end of the neural network to filter the noise. The slew rate limit was set to reject any measurement larger than 7.5 degrees beyond the previous measurement. The purpose of the slew rate limit was to cater for spurious errors, when the eye tracker lost track.

The eye tracker DSP sends the measured eye angles over a RS442 link to a servo controller in the helmet, which commands the mirrors to the appropriate angle. The narrow-field simulator uses this same link. The mirrors are controlled with an analogue servo. The measured step response time of the mirror control system was less than 20ms. The mirror angles have been designed to project the narrow-field by ±10 degrees with no vignetting, and the maximum projected angle is ±15 degrees. The mirrors themselves have a ±7.5 degree rotation range around their own axis.

To ensure that the mirror follows the operator's eye, a calibration is done. The operator is strapped into a jig that holds the helmet fixed, and holds the operator's head in position with a chin rest. The mirror is commanded to known co-ordinates and the operator is asked to fixate on a flashing symbol in the centre of the display, and then to press a button on a mouse. The DSP then records ten eye-tracker raw measurements in succession. This process is repeated for 50 points over a ±15° azimuth and elevation field of regard. The data is uploaded from the DSP onto the narrow field computer that trains a neural network to interpolate between the data. A Levenberg-Marquard back propagation training algorithm is used. The neural network has two input neurons, 10 intermediate neurons and two output neurons. By using the full loop for the calibration, that is from eye in to mirror position out, the mirror was able to follow the operator's eye.

The head tracker is a commercial off-the-shelf ultra-sonic tracker made by Logitech called a 3D mouse. It measures the helmet angle with an accuracy of 0.2 degrees RMS, at a rate of 50 Hz. The helmet position is measured and then an eight-level running average filter is used to filter noise, and to smooth the output.

The simulator uses two computers to generate the narrow and wide field images. The helmet position is transmitted to these computers over a 10 MB Ethernet connection. The narrow field computer measures the mirror angle directly from the RS422 link, using a hardware interrupt controller with a ring buffer. This arrangement ensures that the narrow field computer does not wait for data from the eye tracker DSP before it starts the sight simulation. Copying the appropriate position of the stored scene to the display memory generates the virtual sight pictures. A PC TV converter is used to generate a PAL TV signal from the RGB.. The CRT's are 1" Sony LX3000 viewfinders with a resolution of 400 lines. The typical luminance projected by the CRT's from various parts in the scene has been measured at 70-100 nt. The total latency in the simulator system, from the start of eye movement to the display update, is 60ms.

A photograph of the full system experimental set-up is shown in figure 4.8. The author is wearing the helmet and the computers are, from left to right, the control computer, the narrow field display computer and the wide field display computer. The monitor on the far right shows the eye tracker's image.

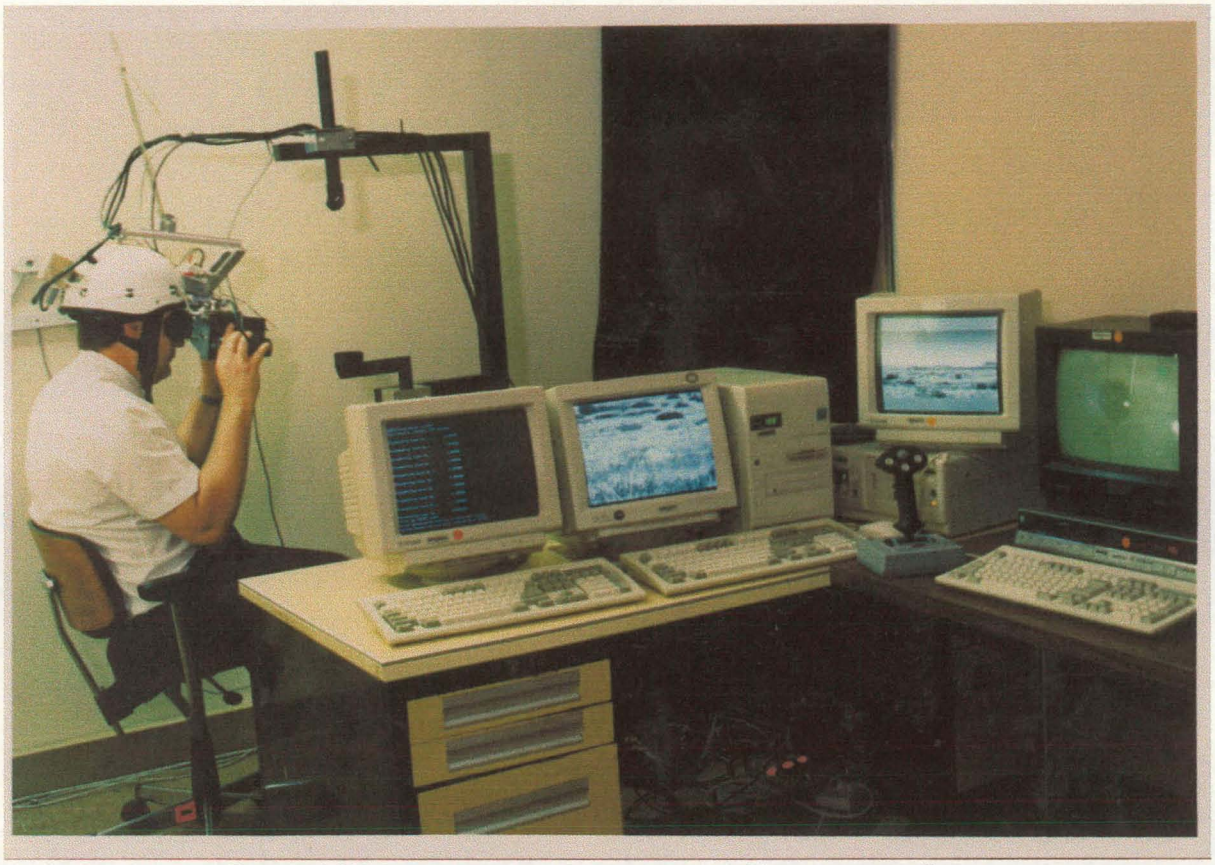


FIGURE 4.8 EYE-SLAVED FOVEAL INSET DISPLAY EXPERIMENTAL SET-UP

4.5.2. Execution of the ESFID Experiment

The eye-slaved display target acquisition experiment was done according to the same protocol as the direct vision and terminal display target acquisition experiments. A total of 20 subjects were required, 10 for each contrast level. The only difference was that the subjects required 20/20 vision without eyeglasses, as the apparatus does not have the capacity to do diopter adjustments.

The apparatus was first calibrated to each operator's eye characteristics to ensure that the eye tracker followed the eye consistently and accurately. The operator then took some time to familiarise himself with the apparatus until he reached a level of consistency in his target acquisition performance. Finally the subjects executed the

experimental protocol, as defined for direct vision. During the experiments the system recorded eye and head movements to determine fixations and saccades.

The tests were done in three sessions for each subject. The first session involved the calibration of the eye tracker. The second session was the training session, where the subjects were familiarised with the helmet, foveal-peripheral alignment and fusion, as well as the scene, typical targets and the search strategy. On average, each subject received about 30 minutes training using four training targets. The reason why the ESFID training took much longer than the other systems, was that most of the training went into teaching the subject not to try and control the narrow field insert with his eye, but to just look at the picture. The operator has a natural tendency to look at the centre of the narrow-field display. As the narrow-field display is rarely positioned at the exact centre of fixation, looking at the centre of the display causes a positive feedback loop. As the eye moves to the centre of the display, the display moves away, etc. As long as the operator concentrates on the picture itself, and fixates on features in the picture there is no problem. The final session involved the actual testing against the sixteen targets, which took another 30 minutes on average. The subjects had no previous exposure to the eye-slaved helmet system, and were not used for the previous tests.

Each run started exactly in the middle of the scene, with the crosshairs on the horizon. All subjects were trained to first immediately move to the left/right and to do a quick scan for the short-range target (within ten seconds), followed by a more leisurely scan back over the rest of the scene (20 seconds). Subjects were encouraged to look around the scene in an as natural as possible manner, which meant moving the head in little steps, and then doing the search with the eyes.

4.5.3. Eye-slaved foveal inset display system Results

The results are summarised in tables 4.6 and 4.7 for the 20% and 30% contrast targets respectively. The subjects gave the following oral feedback:

- The eyepiece was hurting most operators, as it was not ergonomically designed.
- For some subjects, their eyelashes got in the way of the eye tracker, causing the narrow-field display to jump noisily around and confuse the operators. The operators were trained to open their eyes wide whenever the display movement became erratic.

- If the operator looked outside the measurement volume of the helmet tracker, the picture jumped up or down into the limits, which also confused some of the operators. The operators were trained to recognise the limits of the tracker in the scene, and not to search beyond these limits.
- Some of the subjects (72 and 43) experienced a "lazy head" phenomenon, that is, their eyes led their head movements by such an extent that the eye display moved into its limits, and they lost the eye-slaved display movements. This problem could be ascribed to the limited angular projection of 10 degrees in the prototype display. These operators were trained to consciously move their heads so that their eyes would be more centred on the display.
- A general comment from the subjects was that the display was very heavy, and it felt as if they were looking through binoculars, which limited their ability to look around the scene in a natural manner.
- Once the subjects learned to use the ESFID system, their general impression was that of looking at a scene with high-resolution details being filled in wherever they looked. None of the subjects reported a feeling of telepresence. The monochrome images and the relative low resolution used in the experimental hardware did not seem to support the level of realism required for telepresence to occur.

Subject No.	RUN NUMBER																No. of Targets Acquired
	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Quadrant	3	1	4	3	4	2	1	1	4	2	4	2	3	2	3	1	
Range	500	750	250	750	500	1000	250	500	1000	250	750	500	1000	750	250	1000	
8	24.5		3.2	27.6	7.1	21.4	9.9	15.0				20.6					8
14			3.1							9.9		21.8		21.8	5.8		5
22	18.0		9.9				8.0	25.9		7.3	16.0			29.9			7
92	21.0						9.0	13.9		7.6		28.1					5
131	25.0				14.8		9.9			9.5		25.8					5
171	21.8		9.8					25.4		5.0		24.2		25.1			6
55								19.0		5.0		25.6		25.6			4
153	22.3			18.6										29.0			3
18			10.0									22.5	25.0	29.5			4
161	17.1		10.0				5.3	28.1		3.2		23.5		25.6	10.0		8
Mean	21.4	-	7.7	23.1	10.9	21.4	8.4	21.2	-	6.8	16.0	24.0	25.0	26.6	7.9	-	5.5
Std. Dev.	3.0	-	3.5	6.4	5.5	-	1.9	6.1	-	2.5	-	2.4	-	3.0	3.0	-	1.7
Pacq	0.7	0.0	0.6	0.2	0.2	0.1	0.5	0.6	0.0	0.7	0.1	0.8	0.1	0.7	0.2	0.0	0.344

Table 4.6 Eye-slaved foveal inset display system Results for a 20% Contrast Target

Subject No.	RUN NUMBER																No. of Targets Acquired
	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Quadrant	3	1	4	3	4	2	1	1	4	2	4	2	3	2	3	1	
Range	500	750	250	750	500	1000	250	500	1000	250	750	500	1000	750	250	1000	
169	20.5	29.0					6.0			9.9					25.0		5
15	11.0			14.7			9.0	28.1		9.0		25.9		21.6			7
82	25.3			21.8								26.3					3
72	16.6		10.0	17.4	10.0	28.8	5.1	25.6		4.5	11.7	25.7	15.9	21.8			12
62	20.5			29.9			8.0		20.5	25.3			29.9				6
51				19.8			8.0			6.2		7.7		29.9	2.7		6
52	3.2		9.0		12.2	25.7		28.4				22.4		23.0			7
59	19.1	29.7		14.8	14.5		3.6		27.9					27.8	9.2		8
133	15.5			20.0								9.0		29.9			4
43	13.3					29.9	7.3	26.9		2.7		29.9		25.1	8.0		8
Mean	16.1	29.4	9.5	19.8	12.2	28.1	6.7	27.3	24.2	9.6	11.7	21.0	23.0	25.6	11.2	-	6.6
Std. Dev.	6.5	0.5	0.7	5.2	2.3	2.2	1.9	1.3	5.2	8.2	-	9.9	10.0	3.7	9.6	-	2.5
Pacq	0.9	0.2	0.2	0.7	0.3	0.3	0.7	0.4	0.2	0.6	0.1	0.7	0.2	0.7	0.4	0.0	0.413

Table 4.7 Eye-slaved foveal inset display system Results for a 30% Contrast Target

4.6. COMPARATIVE RESULTS

The results for all three tests are plotted together to facilitate comparison between the experiments. The results are plotted against the three independent parameters of target contrast, the target size (range), and the target position (quadrant).

The first plots gives the target acquisition probability, calculated as the ratio of subjects that had successfully acquired a given target in the allocated time over the total number of subjects that searched for that target. The target acquisition probability as a function of target range is shown in figures 4.9 and 4.10, for the 20% and 30% contrast targets respectively. It is clear from these plots that the direct vision performance is consistently better than the other two methods. The terminal display performance is consistently poorer than the eye-slaved display. It is only for 30% contrast targets that the eye-slaved display performance is close to that of direct vision.

Plotting target acquisition performance vs. the target position gives the results shown in figures 4.11 and 4.12. It is clear that the target acquisition performance is best in the second quadrant, and worst in the forth quadrant for all of the display systems, and the target contrasts. Again the direct vision performance is consistently better than the other two methods. The terminal display performance is consistently poorer than the eye-slaved display. It is only for 30% contrast targets in the third and fourth quadrant that the eye-slaved display performance is approaching that of direct vision.

The second performance measure that was used was the mean time it took to acquire the target, if it was acquired. These times are plotted as a function of range in figures 4.13 and 4.14 for target contrasts of 20% and 30% respectively. The error bars indicate the standard deviation in search time. It can be said in general that up to 750m range, the further away the target is, the longer the acquisition time. There is also a clear trend that the ESFID took longer to acquire the targets than any other method.

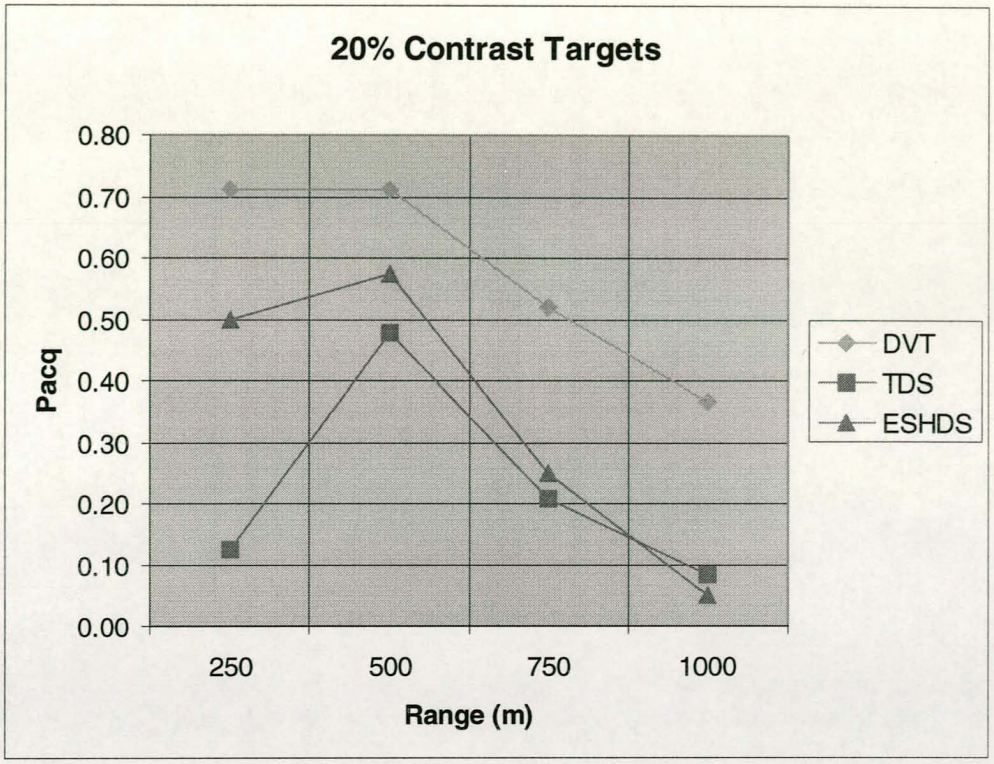


FIGURE 4.9 20% CONTRAST TARGET ACQUISITION PERFORMANCE VS. TARGET RANGE

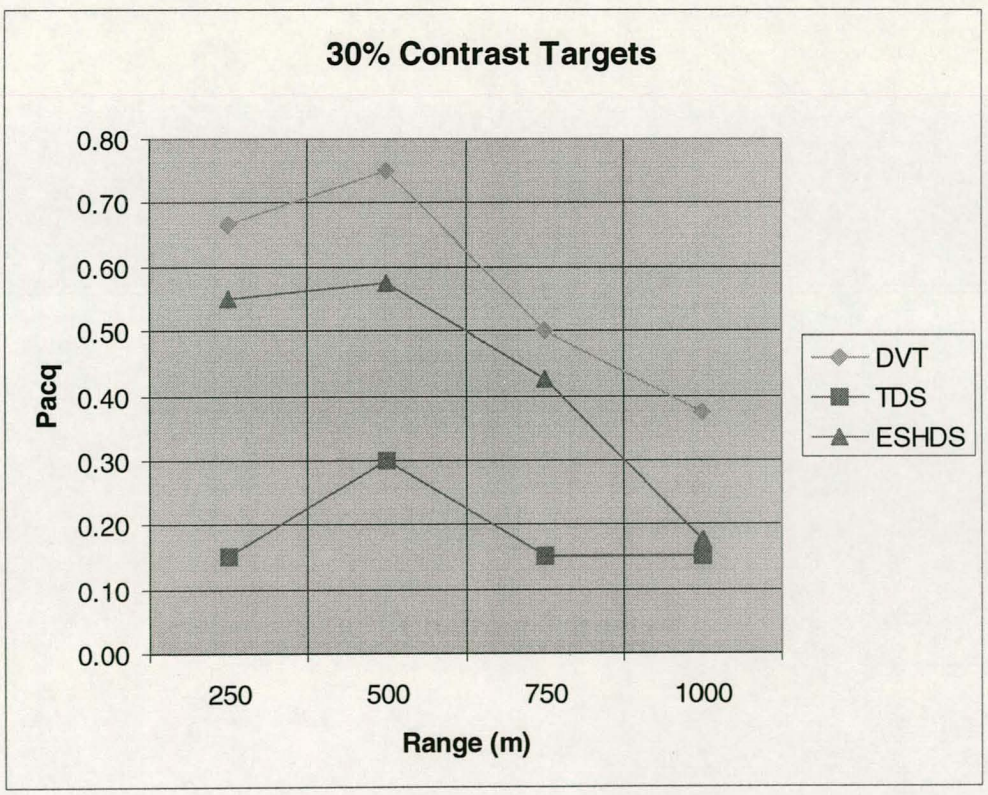


FIGURE 4.10 30% CONTRAST TARGET ACQUISITION PERFORMANCE VS. TARGET RANGE

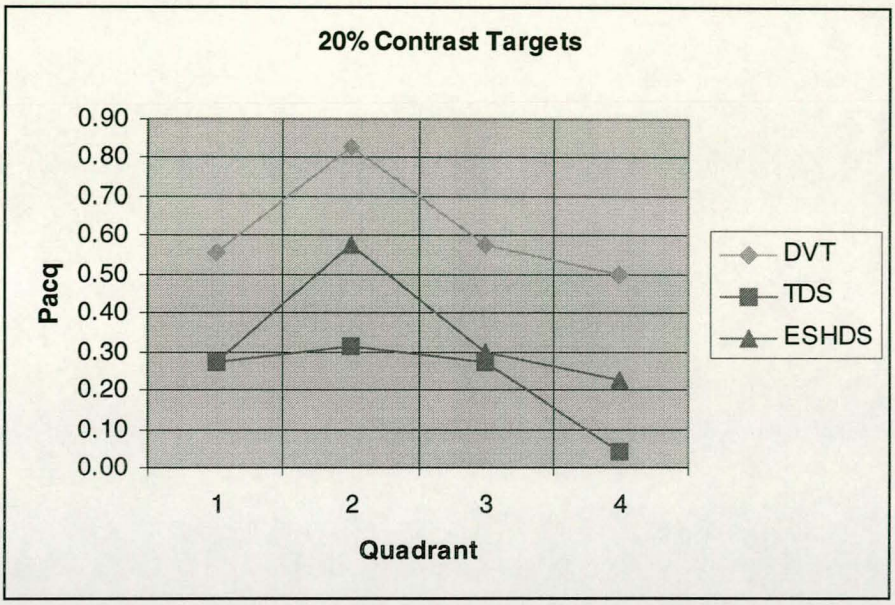


FIGURE 4.11 20% CONTRAST TARGET ACQUISITION PROBABILITY VS. QUADRANT AND EXPERIMENT

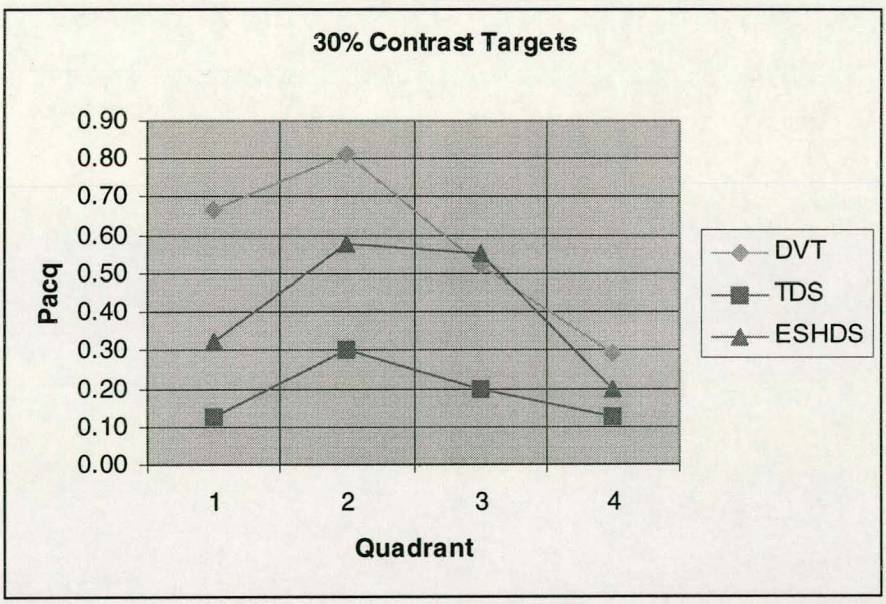


FIGURE 4.12 30% CONTRAST TARGET ACQUISITION PROBABILITY VS. QUADRANT AND EXPERIMENT

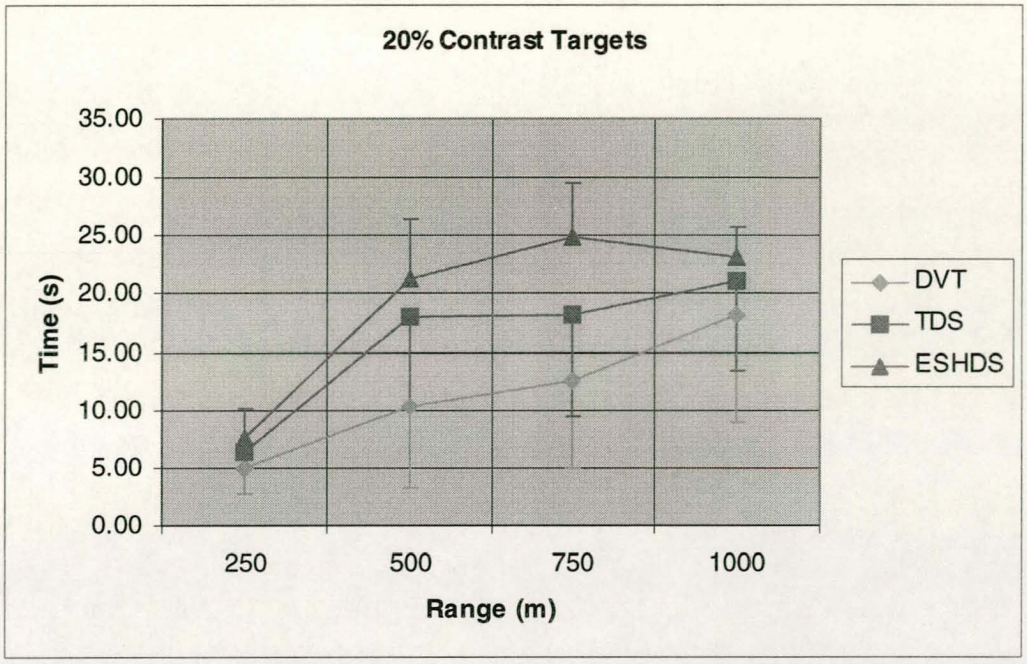


FIGURE 4.13 20% CONTRAST MEAN TARGET ACQUISITION TIME

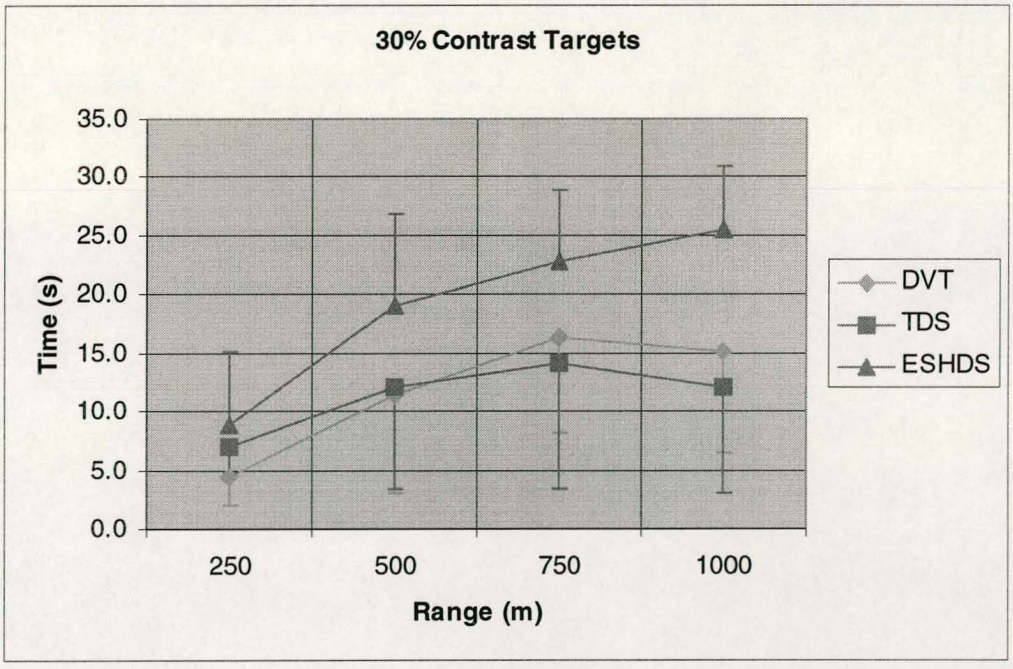


FIGURE 4.14 30% CONTRAST MEAN TARGET ACQUISITION TIME

CHAPTER 5 DISCUSSION

5.1. ANALYSIS OF RESULTS

The purpose of this section is to analyse the results, make observations, give explanations and validate the results. The following questions arise from the results given in Chapter 4:

- What is the effect of the search times that were allocated?
- Why did the TDS and ESFID give lower performance than DVT? A detailed comparative performance analysis will be done.
- Why did the ESFID take longer to acquire targets than any of the other experiments?
- Why is there a consistent difference in target acquisition probability over the different quadrants?
- Do these results confirm the hypotheses of chapter 3.1?

The rest of this paragraph gives in-depth answers to these questions, to check the validity of the results.

5.1.1. Allocated Search Time Analysis

The allocated search time enabled the test subjects in the direct vision experiment to search the whole scene through once. The purpose of the target acquisition experiment is to compare various display methods with one another and if a specific display concept causes longer search times, it is an evaluation result. The search times are therefore only analysed for the direct vision experiment. Intuitively, and from the mathematical model of target acquisition we can show that the detection probability will tend to unity as time tends to infinity. From equations (A.15) and (A.16) in Appendix A we have:

$$\begin{aligned} \lim_{t\to\infty} P_{\det} &= \lim_{t\to\infty}\left(1-\left(1-P_g\right)^{\frac{t}{t_g}}\right) \\ &= 1 \end{aligned} \tag{5.1}$$

for values of $0<P_g\leq 1$ and $t_g>0$. This means that as the search time increases the detection probability approaches unity, even if the probability of detecting the target with a single glimpse P_g is exceedingly low. For this reason, the allocated search time had to be selected arbitrarily, but held constant for each of the different displays.

Figure 5.1 and 5.2 shows histograms of the acquisition times of successful 30s and 10 s maximum search time targets respectively. The 30s maximum search time targets gave a mean time to acquire of 13.29 s and the standard deviation is 8.4 s, while the 10 s targets gave 4.75 s mean and a 2.4 s standard deviation.

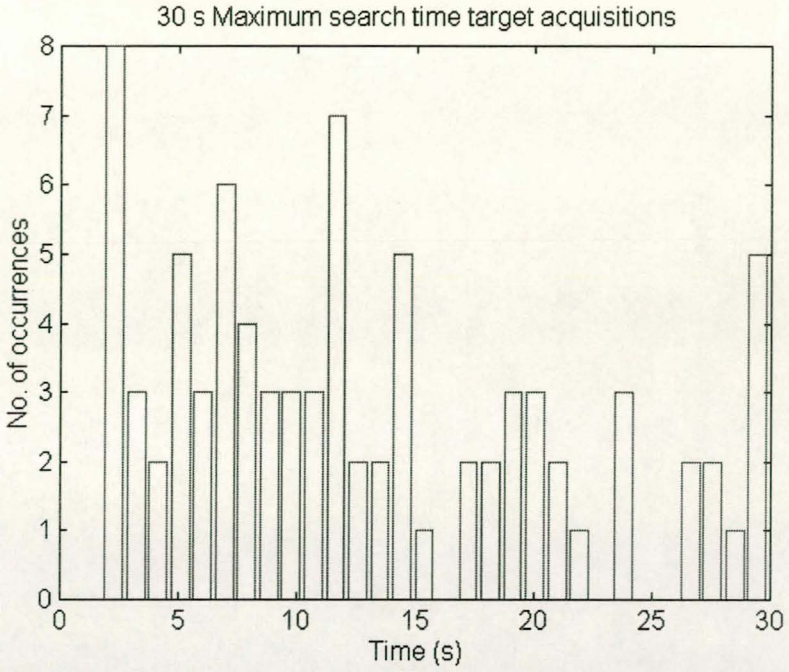


FIGURE 5.1 DISTRIBUTION OF TARGET ACQUISITION TIME FOR 30 S MAXIMUM SEARCH TIME TARGETS

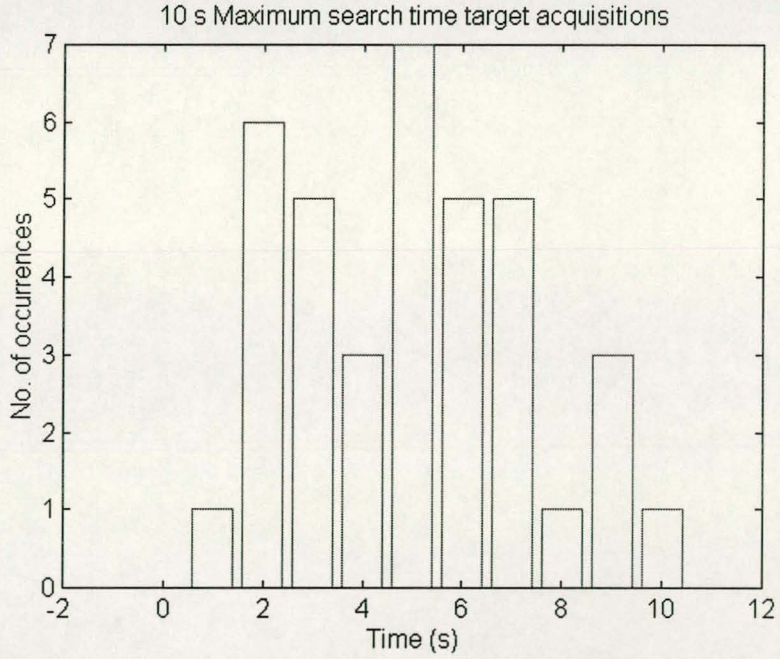


FIGURE 5.2 DISTRIBUTION OF TARGET ACQUISITION TIME FOR 10 S MAXIMUM SEARCH TIME TARGETS

The mean and standard deviation values on tables 4.2 to 4.7 show that for the targets that were acquired, the average acquisition times were quite comfortably inside the maximum allocated search time limit.

5.1.2. Comparative Performance Analysis

The purpose of this section is to analyse the difference of target acquisition performance relative to direct vision. The target acquisition performance results can be classified as follows:

a) Successful target acquisition

b) False target acquisition

c) No target acquisition:

i) The target was never on the display screen

ii) The target was on the display screen, but was not detected

In the case of the direct vision tester, the target was always on the display screen. The display co-ordinates were recorded while the subjects were searching for the target for both the terminal display and the eye-slaved display systems. All the terminal display recorded data, that is sixteen target searches of each of the 22 subjects, were analysed. The narrow-field recording failed for four of the subjects during the eye-slaved display tests, but sixteen good datasets were captured and analysed. From these recordings it is possible to allocate numbers to each of the above categories. The results are summarised in table 5.1.

DISPLAY TYPE	% SUCCESSFUL ACQUISITIONS	% FALSE TARGETS	% UNSUCCESSFUL TARGET ACQUISITIONS	
			Target Never on Display	Target on Display, but not Seen
DVT	57.5	0.0	0.0	42.5
TDS	20.17	2.84	14.49	62.5
ESFID	37.9	0.0	3.12	58.9

Table 5.1 Target Acquisition Analysis

The ESFID gave a 11.37% reduction in the number of targets that were never on the display compared to the TDS. The ESFID search thus covered a much larger area of the scene than the TDS. This ESFID result can be ascribed to the fact that the eye-slaved display has a much higher steering rate, and that the wide-field display gives a reference area for the narrow-field to search in, ensuring that there are fewer gaps in the scene. A large

amount of the performance difference between TDS and ESFID can be explained by the fact that in the TDS some of the targets were simply never on the screen.

Sixty percent of the false target acquisitions for the terminal display can be traced to a single subject, no. 156 (See table 4.4). This specific subject seemed to be unable to discriminate between clutter in the background and targets, and the results of subject 156 should be regarded as an anomaly. False target detections were in general very low which indicates that the targets, if seen, are easily recognisable.

The data was further analysed in an attempt to find an explanation for the situation where the target was on the display screen at some stage, but that the operator did not see it. It was hypothesised that the reason for not seeing the target that is present on the display has a relationship with the search speed, and the minimum distance to the target during the search, that is, the operator may be searching only the centre of the display. Figure 5.3 and 5.4 shows a histogram of the pixel speeds at the moment the display centre moves closest past the target on the screen, for the TDS and the ESFID respectively.

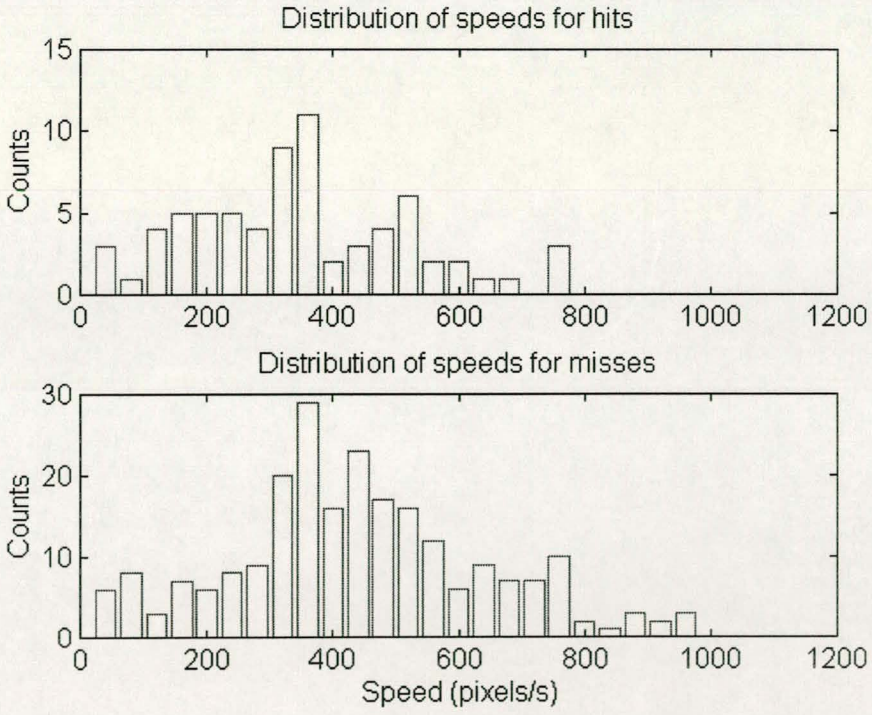


FIGURE 5.3 TDS SEARCH SPEED ANALYSIS

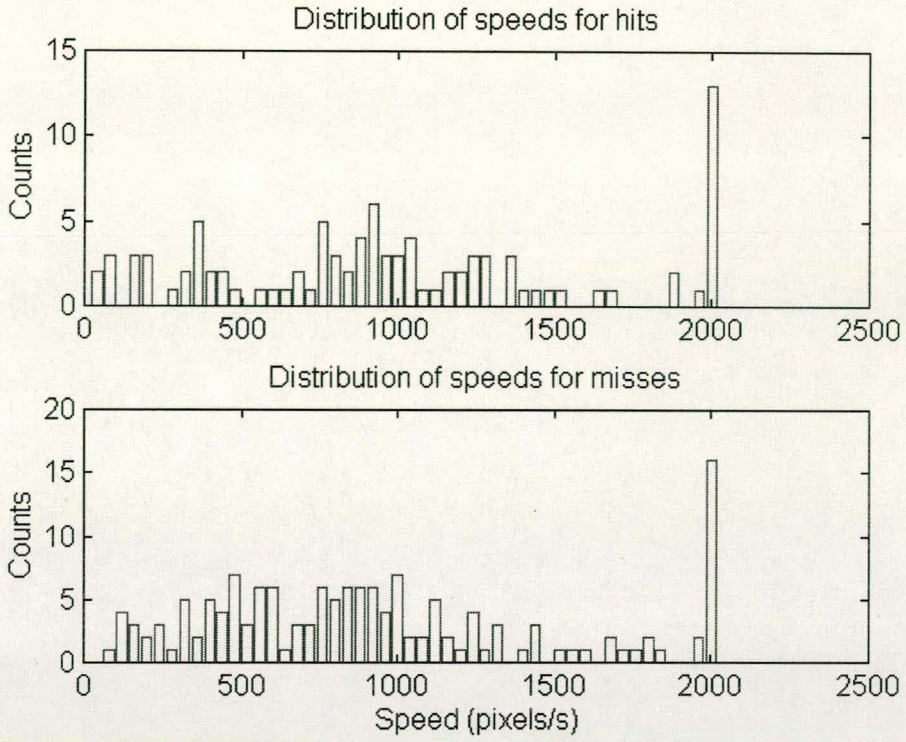


FIGURE 5.4 ESFID SEARCH SPEED ANALYSIS

The mean pixel speed for the TDS system for successful target acquisitions (hits) was 348 pixels/s (4.86°/s), whereas the mean speed for unsuccessful target acquisitions (misses) was 437 pixels/s (6.11°/s). The display speed can be converted to angular rate by multiplying the pixel per second number with the display scale (0.244 mrad/pixel or 0.84arcmin/pixel). It is clear from the histograms that the TDS speed distribution for misses is offset to higher speeds compared to the hits. The mean pixel speed for the ESFID system for successful target acquisitions was 1170 pixels/s (16.36°/s), whereas the mean speed for unsuccessful target acquisitions was 1164 pixels/s (16.28°/s). The histograms show that the ESFID speed distribution for misses is very similar to the hits. The mean search speed over all targets and subjects for the TDS system was 378 pixels/s (5.28°/s), while for the ESFID it was 955 pixels/s (13.35°/s). The higher average search speed of the ESFID can be directly ascribed to the inclusion of eye-motion, and it allows for a larger area of the scene to be examined in a given search time.

The results indicate that search speed is a statistically significant factor for the terminal display system performance, but that it has little effect on the ESFID performance. To explain this result, it is necessary to understand that the main difference between the ESFID and the TDS system lies in the fact that in the TDS the scene moves relative to the observer, and the observer must fixate on a moving image. We have a situation where the target acquisition is dynamic. In the case of the ESFID, the subject uses his natural head and eye

movements and fixates on stationary points, so the visual sensitivity employed is static. Charles Scialfa et al [111] showed that the visual sensitivity peaks at much lower spatial frequencies for dynamic than for static pictures (1.86 vs. 5 cycles per degree). Miller [112] found that visual acuity decrease rapidly with moving targets. His main result is reproduced in figure 5.5 where it is clear that in the case of the TDS, the dynamic visual acuity was in the region of 2 minutes of arc, which is a factor 2 lower than static acuity.

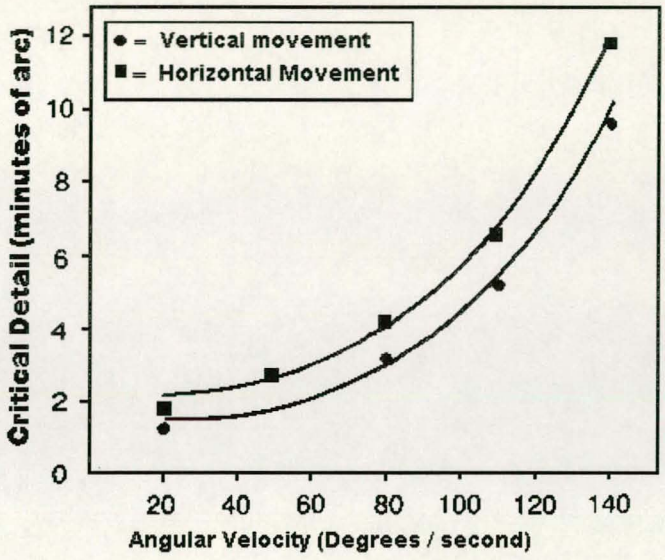


FIGURE 5.5 DYNAMIC VISUAL ACUITY BY MILLER [112]

The other factor that was analysed was the miss distances to the target during the search. Figure 5.6 and 5.7 shows the distribution of minimum distance between the target and the centre of the display screen during search for the TDS and the ESFID systems respectively. The mean miss distance for successful target acquisitions for the TDS system is 72 pixels, while the mean miss distance for misses is 311 pixels. For the ESFID, the mean miss distance is 193 pixels for hits and 241 pixels for misses.

The miss distance results indicate that the TDS require a much smaller miss distance to be acquired than the ESFID. The mean miss distance of 72 pixels is only one degree and the maximum angle at which targets were seen is only 2.5 degrees, all of which indicate that the subjects tend to look only to the middle of the TDS display during search, and that saccadic movements were repressed

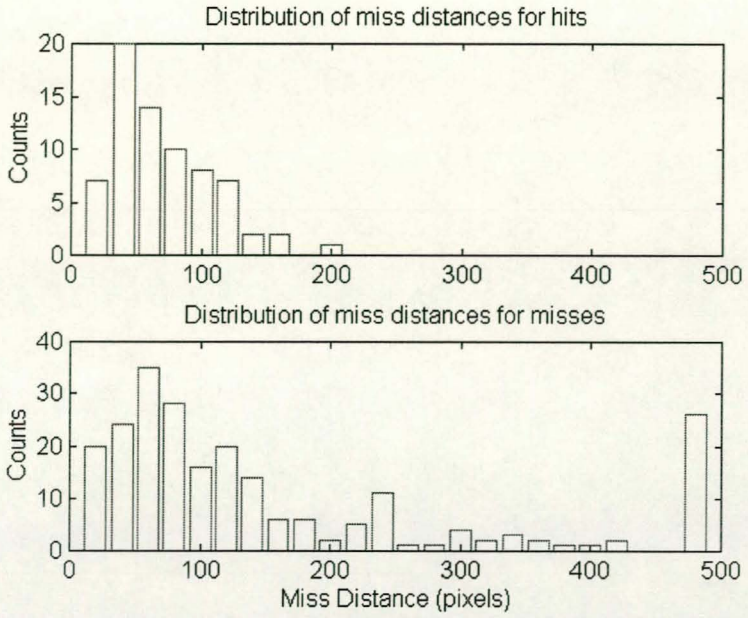


FIGURE 5.6 TDS MISS DISTANCE ANALYSIS

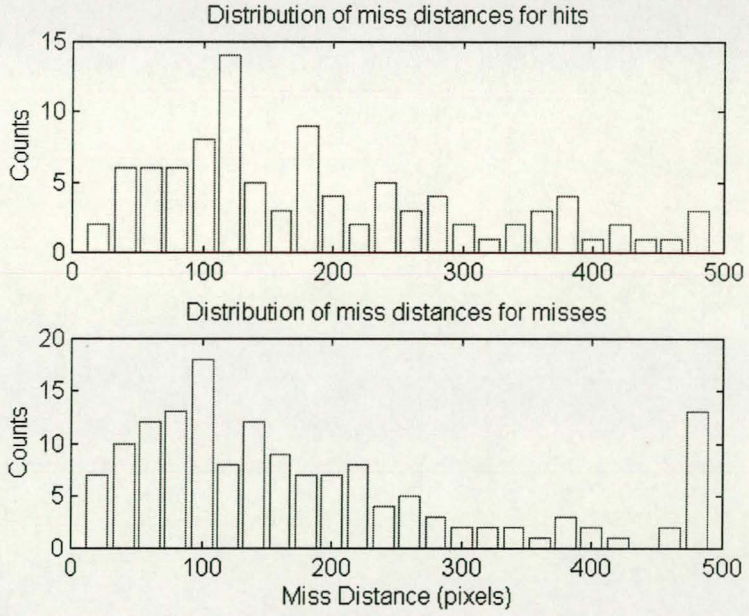


FIGURE 5.7 ESFID MISS DISTANCE ANALYSIS

In the case of the ESFID, the miss distances were much bigger, and quite a number of targets were acquired outside the narrow field of view, in the peripheral display (The narrow field display gives a maximum miss distance of 400 pixels from the centre). Peripheral acquisition did occur in quite a number of cases for the short-range targets that spanned large areas. The fact that there is little difference between hit and miss distributions indicates that the ESFID display supports more natural saccadic eye movement and that the peripheral display also contributed to target acquisition.

5.1.3. ESFID Search Time Analysis

Figures 4.5 and 4.6 shows clearly that the ESFID consistently took longer to acquire a target than the other display systems. The main reason is that in the case of the ESFID the subjects were trained in the use of the apparatus through a common search strategy, while the subjects were allowed to develop their own search strategies for the other display systems. As long as the area to be searched is covered by the search strategy inside the allocated search time, the target acquisition results are independent of search strategy, as was described by Megaw [79].

The typical azimuth search pattern used in the ESFID is shown in figure 5.8, together with a typical scan pattern for the TDS. The ESFID operators were trained to do a quick ten seconds scan followed by a more leisurely 20s search over the 120 degree field of view, to ensure that the full scene was searched. The ESFID operators each had an effective ten seconds added to their search, and an analysis of the search times shows that the mean search time for the TDS for a successful acquisition was 13.6s, compared to the ESFID mean period of 17.7s, a 4.1s longer search time on average. Note the high speed eye movements superimposed on the ESFID track, in figure 3.7.

We conclude that the longer search times associated with the ESFID can be ascribed to the trained search pattern, whereas in the TDS and DVT the subjects were free to choose their own search strategy.

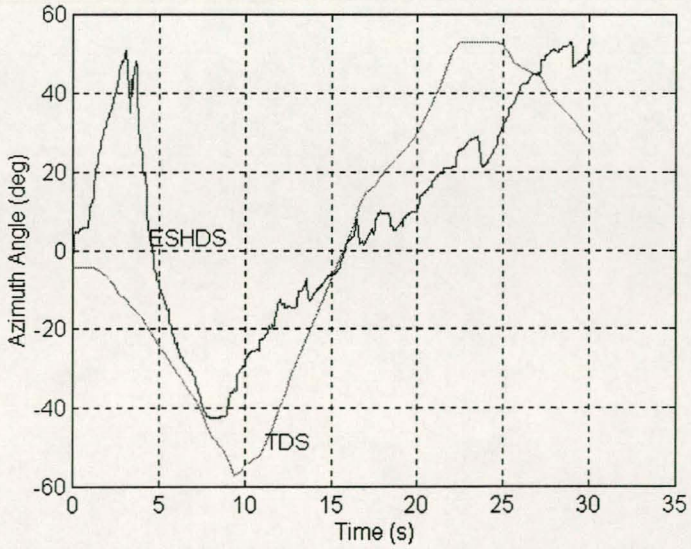


FIGURE 5.8 TYPICAL TDS AND ESFID AZIMUTH SEARCH PATTERNS

5.1.4. Target Acquisition Probability vs. Quadrants

The results shown in figures 4.3 to 4.4 indicate that the target acquisition probability is consistently worse in the fourth quadrant, and better in the second quadrant, for all of the display systems and contrast levels. The possible reasons for this were postulated as follows:

- The subjects did not search the fourth quadrant as much as the other quadrants, and spent more time in the second quadrant.
- The fourth quadrant has a much higher background clutter, and the second quadrant has much lower clutter.

The search time analysis determined the percentage time that was spent on average in each quadrant measured over all the subjects and test runs. The result is shown on table 5.2.

Display	Q1	Q2	Q3	Q4
TDS	26.4	27.0	26.1	20.5
ESFID	16.6	30.1	31.6	21.7

Table 5.2 Percentage Time Spent Per Quadrant

From table 5.2 we see that the fourth quadrant has indeed been searched less than the others, except for the first quadrant of the ESFID. The second quadrant does not show a significantly higher search time than the others. The fact that a higher target acquisition probability was achieved by the ESFID in the first quadrant than the fourth quadrant, but with much less search time, indicates that the search time is not a very significant factor.

To analyse the effect of clutter, Wooton's [101] model was used to obtain the clutter values. Although Waldman [102] mentioned that the clutter factor determined by this technique does not match his model accurately, Waldman did find that Wootton's clutter measure gave good relative values for clutter.

The clutter value was determined for each of the target positions, i.e. for each quadrant and target range. The clutter value was determined in the horizontal and vertical direction, and the average value between the two was used. The clutter model results for the scene in figure 3.2 are given in table 5.3.

Quadrant	Range [m]	Clutter (Horizontal)	Clutter (Vertical)	Clutter (Average)
1	250	0.133	0.251	0.192
	500	0.525	0.404	0.464
	750	0.34	0.589	0.466
	1 000	0.43	0.475	0.454
2	250	0.288	0.231	0.26
	500	0.197	0.426	0.311
	750	0.313	0.607	0.46
	1 000	1.45	1.2	1.33
3	250	0.131	0.248	0.189
	500	0.573	0.492	0.533
	750	0.367	0.745	0.556
	1 000	0.911	1.907	1.409
4	250	0.22	0.068	0.144
	500	0.39	0.148	0.27
	750	0.67	1.26	0.967
	1 000	1.54	0.983	1.26

Table5.3 Wootton's Clutter Model Values for the test Scene

The 1000m results were not very reliable as the targets were placed in gaps in a solid tree line, as shown in figure 3.2. The clutter values for the 250 to 750 m targets, which made up most of the successful target acquisitions, were averaged and shown on table 5.4.

	Q1	Q2	Q3	Q4
Wootton'sClutter Value	0.265	0.244	0.3	0.326

Table 5.4 Mean Clutter Value per Quadrant (250-750m)

The clutter data in table 5.4 corresponds to the observed acquisition results. The fourth quadrant has a higher clutter value for target ranges up to 750 m than all the other quadrants, and the second quadrant has the lowest level of clutter.

It is concluded that target position does not contribute significantly to the target acquisition probability, there are other factors such as clutter that also influence the result. The target position was discarded as an independent variable for the rest of the analysis in this chapter. The number of independent experimental treatises of each target acquisition factor (Target acquisition system, target/background contrast and target range) is then between 40 and 52 as shown on tables 4.2 to 4.8.

5.1.5. Hypothesis Testing

The assumptions and hypothesis in chapter 3.1 were tested with the experimental results obtained in chapter 4 using statistical analysis techniques. The truth or falsity of a statistical hypothesis is never known with absolute certainty unless the entire population is examined, which is of course impractical. Instead, the population was randomly sampled as described in the experiments in chapter 4. Evidence from the sample that was inconsistent with the stated hypothesis led to the rejection of the hypothesis whereas evidence supporting the hypothesis led to its acceptance. It is important to realise that the acceptance of a statistical hypothesis is the result of insufficient evidence to reject it and does not necessarily imply that it is true. The hypothesis that is tested is termed the null hypothesis and is denoted by H_0. The rejection of H_0 leads to the acceptance of an alternative hypothesis, denoted by H_1. The level of significance is defined as the probability of rejecting the null hypothesis when the null hypothesis is in fact true, and is denoted by the Greek letter α. A small number for α implies a high level of confidence in the final hypothesis.

a) Merging Hypothesis Test

The first assumption was that finite pixel wide-field (e.g. 600 pixels/60°) information and finite pixel narrow-field (e.g. 600 pixels/10°) information could be displayed into separate eyes with the human brain fusing the two pictures into an apparent 6000 pixel display

The merging hypothesis was extensively tested with the two experiments described in chapter 4.1 and 4.2. The first experiment used direct vision to test the hypothesis, while the second experiment used an

electro-optic sensor to extend the merging hypothesis to electro-optic sensing and to determine the alignment and zoom tolerances under which the merging hypothesis holds true. The direct vision fusion experiment used a piece of cardboard with appropriate cut-outs to generate the two visual channels, thus providing full direct vision resolution.

The fusion experiments are of the type known as binomial experiments, as they possessed the following properties [113]:

a) The experiments consisted of n repeated trials.

b) Each trial resulted in an outcome that may be classified as a success (fusion) or a failure (no fusion).

c) The probability of success, denoted by p, remains constant from trial to trial, which is the percentage of the population for which the hypothesis holds true.

d) The repeated trials were independent, by using a different subject for each trial.

Binomial experiments follow binomial statistics, which were used to test the hypothesis.

The direct vision fusion experiment together with the electro-optic fusion experiment gave 27 successes out of 28 trials, for a random sample of the population. As the number of trials was relatively small, the statistical significance of these results was tested using binomial statistics. The direct vision fusion experiment's results may lead to the hypothesis that 85% of the population will achieve fusion. That is, we are testing the null hypothesis:

i) H_0: $p = p_0 = 0.85$,

ii) and the alternative hypothesis H_1: $p > 0.85$.

iii) with a level of significance $\alpha = 0.1$

iv) The test statistic is the binomial variable X with $p = 0.85$ and the sample number $n = 28$.

v) Computation to determine $P=P(X \geq x$ when $p = p_0)$ where the value $x = 27$ is the number of successes in our sample size $n=28$ and $p_0 = 0.85$:

$$\begin{aligned} P &= P(X \geq 27 \text{ when } p = 0.85) \\ &= 1 - P(X < 27 \text{ when } p = 0.85) \\ &= 1 - \sum_{x=0}^{26} b(x;\ 20, 0.85) \\ &= 0.0627 < (\alpha = 0.1) \end{aligned}$$

vi) Decision: Reject H_0 and conclude that the alternative hypothesis is true, i.e. more than 85% of the population can achieve fusion, with a 0.1 level of significance.

The above calculation can be refined and the percentage of the population for which the hypothesis holds true can be calculated as a function of the level of significance:

Level of significance	0.025	0.05	0.1	0.15	0.2
P >	0.816	0.841	0.868	0.884	0.896

Table 5.5 Hypothesis Probability vs. level of significance

After the fusion experiments, subjects were selected to exclude monoscopic fixation, strabismus and low visual acuity. The later ESFID experiment used 20 trials. All 20 of the 20 pre-selected subjects achieved successful fusion. The hypothesis test is done in a similar way than for the fusion experiments:

i) H_0: $p = p_0 = 0.9$,

ii) and the alternative hypothesis H_1: $p > 0.9$.

iii) with a level of significance $\alpha = 0.125$ (one in eight chance of error)

iv) The test statistic is the binomial variable X with $p = 0.9$ and the sample number $n = 20$.

v) Computation to determine $P=P(X \geq x \text{ when } p = p_0)$ where the value $x = 20$ is the number of successes in our sample size $n=20$ and $p_0 = 0.9$:

$$\begin{aligned} P &= P(X \geq 20 \text{ when } p = 0.9) \\ &= 1 - P(X < 20 \text{ when } p = 0.9) \\ &= 1 - \sum_{x=0}^{19} b(x;\ 20, 0.85) \\ &= 0.1216 < (\alpha = 0.125) \end{aligned}$$

vi) Decision: Reject H_0 and conclude that the alternative hypothesis that more than 90% of a pre-selected population can achieve fusion is true, with a 0.125 level of significance.

b) TDS Performance Assumption Test

The assumption is that current technology teleoperated sensors, which consist of a joy-stick controlled television camera and a display, give inferior target acquisition performance against single military targets, compared to direct vision.

The direct vision and the terminal display test experiments that are described in chapter 4.3 and 4.4 were conducted to test the hypothesis. These target acquisition experiments are also of the type known as binomial experiments, as they possess similar properties as the split-eye foveal/peripheral tests.

a) The experiments consisted of n repeated trials.

b) Each trial resulted in an outcome that may be classified as a success (target acquired in the allocated time) or a failure (no target acquired).

c) The probability of success, denoted by p_0, remains constant from trial to trial, that is the target acquisition probability.

d) The repeated trials were independent, by using a different subject, or a different target position for each trial.

Binomial experiments follow binomial statistics, which are used to test the hypothesis. For larger sample numbers, such as in the target acquisition experiments where the number of samples in each trial varied between 40 and 52, the binomial statistics can be approximated with the normal curve. This approximation is accurate as long as the probabilities are not extremely close to 0 or 1. The mean μ and standard deviation σ of the normal curve is given by $\mu = n\, p_0$, and $\sigma = n\, p_0\, (1- p_0)$ respectively.

The testing of Hypothesis 2 concerns the testing of the difference between two proportions, that is the actual target acquisition probability of the direct vision test (p_1) compared to the terminal display test (p_2). The null hypothesis that $p_1 = p_2$ will be tested against the alternative hypothesis that $p_1 \neq p_2$. This is equivalent to testing the null hypothesis that $p_1 - p_2 = 0$ against the alternative hypotheses that $p_1 - p_2 \neq 0$. The statistic is based on the random variable $\hat{p}_1 - \hat{p}_2$. As shown in reference [113], the z value of the normal distribution used for testing $p_1 = p_2$ is given by the following:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

Where $\hat{p}_1$ is the proportion of successful trials x_1 in sample n_1,

$\hat{p}_2$ is the proportion of successful trials x_2 in sample n_2,

$\hat{p}$ is the pooled estimate of the proportion of both samples, $\hat{q} = 1 - \hat{p}$ and:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

To test the alternative hypothesis of $p_1 > p_2$ at the α level of significance, the critical region is $z > z_\alpha$. To demonstrate the test, the 250m range, 20% contrast target data is used, as given by table 4.2 and table 4.4 for the direct vision and terminal display tests. Let p_{DVT} and p_{TDS} be the actual target acquisition probabilities for the DVT and TDS respectively. The procedure is as follows:

i) H_0: $p_{DVT} = p_{TDS}$,

ii) and the alternative hypothesis H_1: $p_{DVT} > p_{TDS}$.

iii) with a level of significance $\alpha = 0.025$

iv) The critical region follows from the normal table as $z > 1.96$.

v) The computations are as follows:

$$\hat{p}_{DVT} = \frac{x_{DVT}}{n_{DVT}} = \frac{37}{52} = 0.7115$$

$$\hat{p}_{TDS} = \frac{x_{TDS}}{n_{TDS}} = \frac{6}{48} = 0.125$$

$$\hat{p} = \frac{x_{DVT} + x_{TDS}}{n_{DVT} + n_{TDS}} = \frac{37 + 6}{52 + 48} = 0.43$$

$$z = \frac{0.7115 - 0.125}{\sqrt{(0.43)(0.57)[(1/52) + (1/48)]}}$$

$$= 5.92 > z_\alpha$$

vi) Decision: Reject H_0 and agree that the direct vision target acquisition probability is higher than the terminal display system's target acquisition probability with a 0.025 level of significance.

The analysis can also be done in reverse, that is, the limit of the level of significance for which H_0 will be rejected can by computed, given the results of table 4.2 to table 4.5. The result of this analysis is shown on table 5.6.

RANGE (m)	20% CONTRAST				30% CONTRAST			
	$\frac{x_{DVT}}{n_{DVT}}$	$\frac{x_{TDS}}{n_{TDS}}$	Critical z	α/2	$\frac{x_{DVT}}{n_{DVT}}$	$\frac{x_{TDS}}{n_{TDS}}$	Critical z	α/2
250	37/52	6/48	5.919	<0.0003	32/48	6/40	4.872	<0.0003
500	37/52	23/48	2.37	0.0089	36/48	12/40	4.221	<0.0003
750	27/52	10/48	3.217	0.0007	24/48	6/40	3.449	0.0003
1000	19/52	4/40	3.348	0.0004	18/48	6/40	4.157	<0.0003

Table 5.6 Level of Significance for which $P_{DVT} > P_{TDS}$

From table 5.6 it can be concluded that the DVT give higher target acquisition probabilities than the TDS with a >99% confidence, that is a less than a 1% chance of a error, for all the targets tested.

c) Hypothesis 1 Test

The thesis hypothesis states that by combining head tracking and eye-tracking in an ESFID system, similar target acquisition performance to direct human vision is possible.

The ESFID target acquisition experiment, described in chapter 4.5, was conducted to test the third hypothesis, together with the other two control experiments, the direct vision and the terminal display experiments. The ESFID target acquisition experiments were binomial, for the same reasons that the other target acquisition experiments are binomial. The ESFID results are first compared to the terminal display results to determine if there is a significant performance increase and then to the direct vision results to determine if the performance is similar.

The null hypothesis is that the ESFID target acquisition probability (p_{ESFID}) is equal to the terminal display target acquisition probability (p_{TDS}). To test the alternative hypothesis of $p_{ESFID} > p_{TDS}$ at the α level of significance, the critical region is $z>z_{\alpha}$. The procedure is as follows:

i) The null hypothesis is H_0: $p_{ESFID} = p_{TDS}$,

ii) and the alternative hypothesis H_1: $p_{ESFID} > p_{TDS}$

iii) with a level of significance calculated from the normal table for the critical region in (iv)

iv) The critical region is calculated as follows, and the results are summarised in table 5.7:

$$\hat{p}_{ESFID} = x_{ESFID} / n_{ESFID}$$

$$\hat{p}_{TDS} = x_{TDS} / n_{TDS}$$

$$\hat{p} = \frac{x_{ESFID} + x_{TDS}}{n_{ESFID} + n_{TDS}}$$

$$z = \frac{\hat{p}_{ESFID} - \hat{p}_{TDS}}{\sqrt{\hat{p}(1-\hat{p})\left[(1/n_{ESFID}) + (1/n_{TDS})\right]}}$$

v) Decision: Reject H_0 and agree that the ESFID target acquisition probability is higher than the terminal display system's target acquisition probability for the 250m 20% contrast target and for 250-750m 30% contrast targets, with a 0.01 level of significance (99% confidence). For the 250-1000m range 20% contrast targets, and the 1000m range 30% contrast targets, there is insufficient evidence to conclude that the ESFID has a different target acquisition probability than the terminal display system.

RANGE (m)	20% CONTRAST				30% CONTRAST			
	$\frac{x_{ESFID}}{n_{ESFID}}$	$\frac{x_{TDS}}{n_{TDS}}$	Critical z	α/2	$\frac{x_{ESFID}}{n_{ESFID}}$	$\frac{x_{TDS}}{n_{TDS}}$	Critical z	α/2
250	20/40	6/48	3.839	<0.0003	22/40	6/40	3.751	<0.0003
500	23/40	23/48	0.896	0.185	23/40	12/40	2.479	0.0066
750	10/40	10/48	0.464	0.321	17/40	6/40	2.717	0.0034
1000	2/40	4/40	-0.617	0.768	7/40	6/40	0.303	0.381

Table 5.7 Level of Significance for which $P_{ESFID} > P_{TDS}$

Finally, the target acquisition probabilities of the ESFID and DVT systems are compared. Let p_{ESFID} and p_{DVT} be the true target acquisition probabilities for the ESFID and DVT systems respectively. The thesis statement is the null hypothesis that the target acquisition probabilities are similar for the two systems. The alternative hypothesis is that they are not the same. The testing procedure is as follows:

i) H_0: $p_{ESFID} = p_{DVT}$,

ii) and the alternative hypothesis H_1: $p_{ESFID} \neq p_{DVT}$

iii) with a level of significance α calculated from the normal table for the critical region $Z > z_{\alpha/2}$ and $Z > -z_{\alpha/2}$ in (iv)

iv) The critical region is calculated as follows, and the results are summarised in table 5.8:

$$\hat{p}_{DVT} = x_{DVT} / n_{DVT}$$

$$\hat{p}_{ESFID} = x_{ESFID} / n_{ESFID}$$

$$\hat{p} = \frac{x_{DVT} + x_{ESFID}}{n_{DVT} + n_{ESFID}}$$

$$z = \frac{\hat{p}_{DVT} - \hat{p}_{ESFID}}{\sqrt{\hat{p}(1-\hat{p})\left[(1/n_{DVT}) + (1/n_{ESFID})\right]}}$$

RANGE (m)	20% CONTRAST				30% CONTRAST			
	$\frac{x_{DVT}}{n_{DVT}}$	$\frac{x_{ESFID}}{n_{ESFID}}$	Critical z	$\alpha/2$	$\frac{x_{DVT}}{n_{DVT}}$	$\frac{x_{ESFID}}{n_{ESFID}}$	Critical z	$\alpha/2$
250	37/52	20/40	2.072	0.019	32/48	22/40	1.119	0.1316
500	37/52	23/40	1.363	0.0864	36/48	23/40	1.739	0.041
750	27/52	10/40	2.611	0.0045	/48	17/40	0.702	0.2414
1000	19/52	2/40	3.573	<0.0003	18/48	7/40	2.072	0.019

Table 5.8 Level of Significance for which $P_{ESFID} = P_{DVT}$

v) Decision: Accept H_0 and agree that there is insufficient evidence to reject the claim that the ESFID target acquisition probability is the same as the DVT for all the targets (except the 750 and 1000m 20% contrast targets), with a 0.04 level of significance.

The ESFID system does have a lower target acquisition probability than direct vision for the 750m and 1000m 20% contrast targets, with a 0.01 level of significance. It is also clear that on average, the ESFID target acquisition probability is lower than direct vision. This result is explained with the aid of a mathematical model in the next section.

5.2. MATHEMATICAL MODEL ANALYSIS

The purpose of the mathematical model is to give a more general prediction of the target acquisition performance, as a function of the various scene parameters.

The target acquisition results are generalised with the aid of a mathematical model of the target acquisition systems. The mathematical model is used to determine the target acquisition performance sensitivity to the design parameters. Recommendations are then made on the system design and the achievable performance.

5.2.1. Mathematical Model Calibration

Waldman's target acquisition and Stark's Markov fixation models were calibrated with the input parameters of the actual target search scene, and the human factor coefficient were adjusted until the model matched the actual acquisition performance. The parameters that have been derived from the scene were:

- Markov fixation model
- Scene Parameters such as light levels, target size and sky background ratio
- Clutter Factor

a) Markov Fixation Model

The fixation model for the ESFID is derived from the recorded eye position coordinates of all of the ESFID subjects and for all of the targets. Sixteen subjects each with 16 targets were used for a total of 183 578 number of recorded points. The coordinates for each recording interval were binned on a two dimensional histogram to determine the fixation probability density function over the scene. The bin size that gave the best definition of the fixation peaks was found to be 17.5 pixels by 17.5 pixels on the narrow-field scene. The fixation histogram is shown in figure 5.9.

A peak detector was used to determine the centre of the fixations in the fixation histogram. A total of 469 fixation peaks were detected, and they are shown in figure 5.10. These fixation points were ordered using ascending azimuth coordinates. The probability that each fixation point was occupied is known as the Markov $\mathbf{M_0}$ matrix. This matrix is shown in figure 5.11 as a percentage, not a probability.

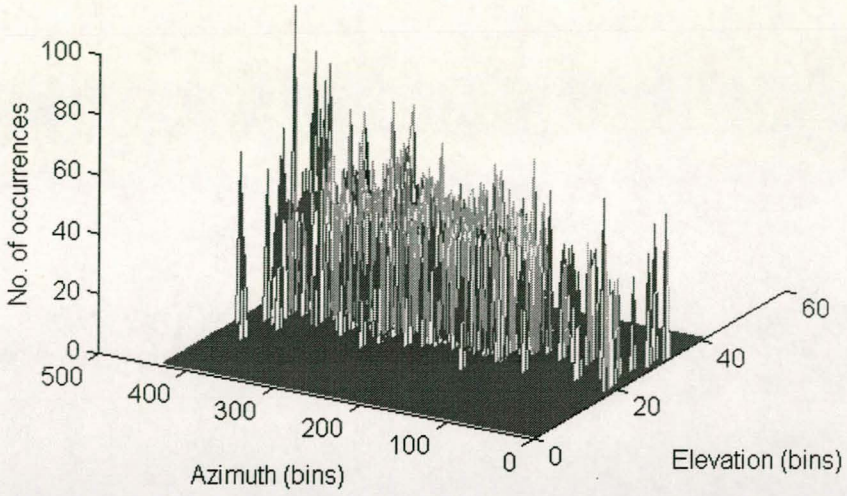


FIGURE 5.9 FIXATION HISTOGRAM

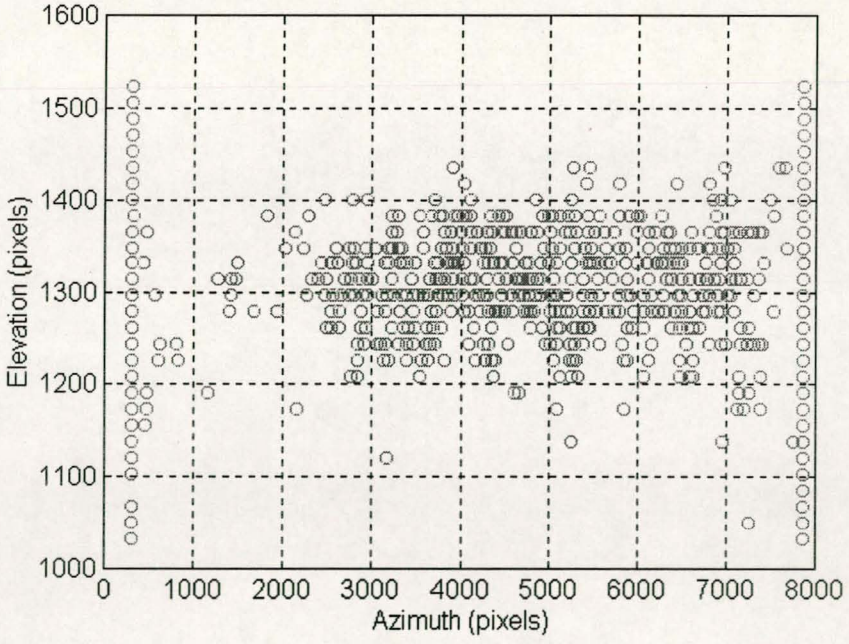


FIGURE 5.10 FIXATION PEAKS

The transition probability from each of these peaks to any of the other peaks was determined to get the $\mathbf{M_1}$ Markov transition matrix. The number of switches from a fixation point **i** to the next fixation point **j**

was measured. The result was stored in a 469x469 intermediary matrix **A** in which the elements $\mathbf{A}_{ij}$ were the counts of the number of switches from fixation **i** to fixation **j**. By doing a line by line normalisation, the $\mathbf{M}_1$ Markov transition matrix was determined. The elements of the $\mathbf{M}_1$ Markov matrix are shown graphically in figure 5.12. The $\mathbf{M}_1$ matrix is diagonally dominant because the sequence in fixations was limited to nearby points in azimuth.

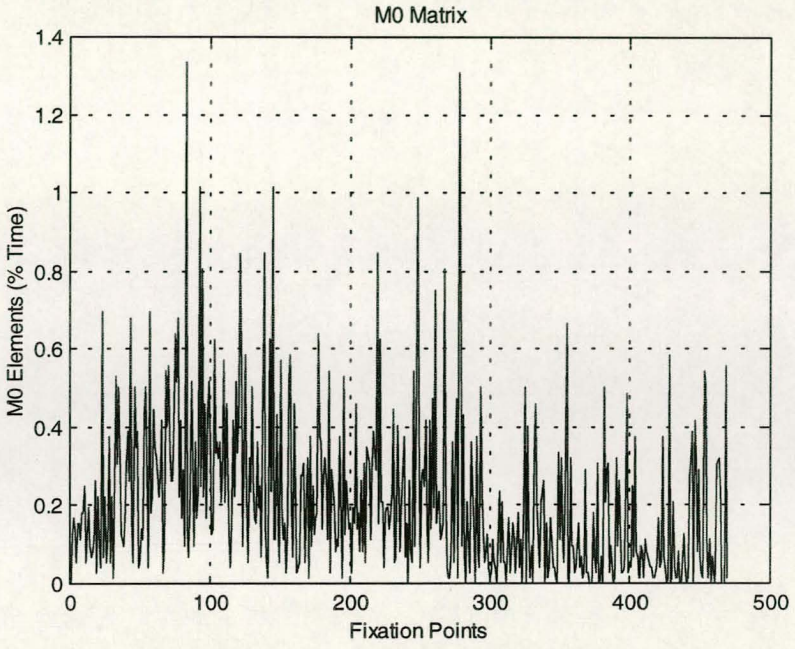


FIGURE 5.11 MARKOV M0 MATRIX

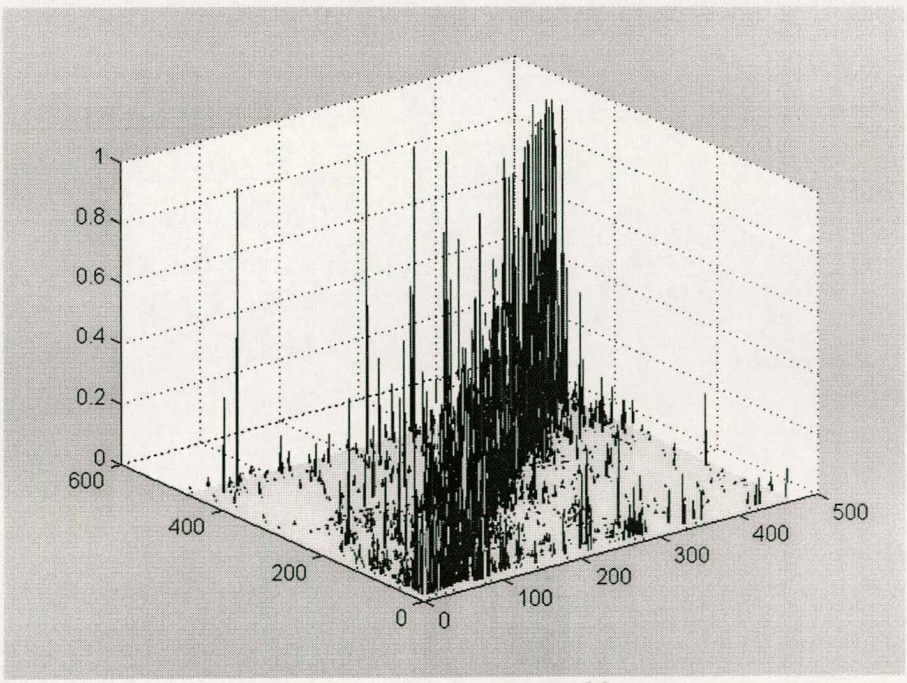


FIGURE 5.12 MARKOV $\mathbf{M}_1$ TRANSITION MATRIX

The Markov model was used to generate the sequence of fixation points as determined by the peak detector in the mathematical model. The model always starts with the same fixation point, point no 193 which is in the middle of the scene with coordinates (4095,1277). The next fixation points follows from row 193 in the $\mathbf{M_1}$ matrix. A random number **r** is drawn with a uniform distribution between zero and one. The coordinate of the element in the 193rd row where the cumulative probability along the row exceeds **r** gives the number of the next fixation point. The coordinates of that fixation point follows from the mapping of fixation points against coordinates as shown in figure 5.10. The time of the fixation is modelled according to equation (A.9) of Waldman's model as given in Appendix A, after which the above process is repeated to get the next fixation point.

b) Scene Parameters

The scene parameters that were measured during the experiment are the following:

STime=30.0	% Search time for 500m-1000m ranges [s]
STime=10.0	% Search time for 250m ranges [s]
LOB=100	% Background luminance [fL]
Lmon=75	% Monitor maximum luminance [fL]
S=1.5	% Sky-ground luminance ratio
TgtSize=4.55	% Average target dimension [2.1 x 7.0m]
C0=0.2 and 0.3	% Scene target background contrast
Saz=120.;Sel=12.	% Search area, azimuth and elevation [deg]
Rvis=20000.0	% Visibility range [m]
BIC=1.0	% Battlefield induced contaminants factor
Ptype=1	% Precipitation type 1-rain, 2-sleet, 3-snow
r=0.0	% Precipitation rate [mm/h]
ERes=0.0003	% Eye limiting resolution [rad]

c) Clutter Factor

The first parameter that is determined from the scene is the clutter factor. Wootton's co-occurrence matrix technique [101] as given in table 5.2 gives good relative values but does not give accurate absolute target acquisition probabilities. As this dissertation is a comparative study, it was decided to use the direct vision (the control experiment) results to determine the best clutter factor for each target to match the results obtained.

The model of Waldman was implemented with the above mentioned scene parameters, and for direct vision. This model was implemented in MATLAB™ as VDDVT.m and the listing is given in appendix B. The predicted target acquisition probabilities as a function of clutter values were determined, and plotted in figure 5.13 and 5.14 for the 20% and 30% contrast targets, and for the various ranges and search times used in the target acquisition experiments. The clutter value that gives the best approximation to the measured target acquisition result that was found in the experiment was then read off these graphs. The resulting clutter values are given in table 5.9

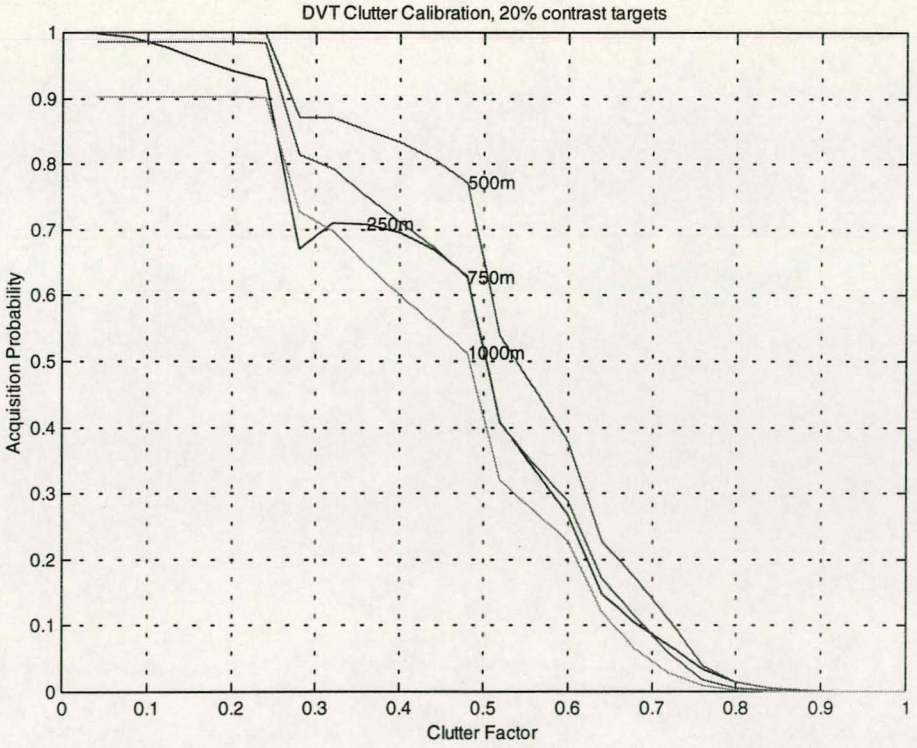


FIGURE 5.13 DVT CLUTTER CALIBRATION CURVES FOR 20% CONTRAST TARGETS

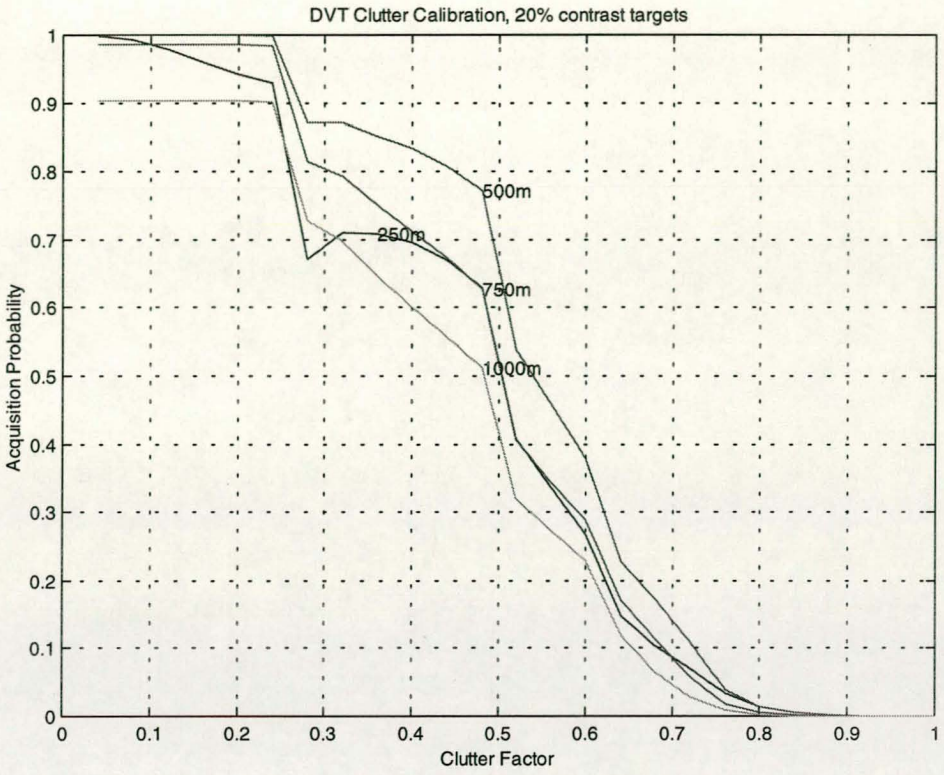


FIGURE 5.14 DVT CLUTTER CALIBRATION CURVES FOR 30% CONTRAST TARGETS

Quadrant	Range [m]	20% Contrast		30 % Contrast	
		Pacq	Clutter	*Pacq*	Clutter
1	250	0.77	0.2648	0.83	0.4501
	500	0.77	0.4802	1.0	0.2680
	750	0.31	0.5846	0.42	0.5975
	1 000	0.38	0.5076	0.42	0.5231
2	250	0.85	0.2524	1.0	0.2420
	500	0.92	0.2648	0.83	0.5066
	750	1.0	0.2434	1.0	0.2457
	1 000	0.54	0.4553	0.42	0.5231
3	250	0.77	0.2648	0.58	0.5251
	500	0.77	0.4802	0.83	0.5066
	750	0.38	0.5368	0.25	0.6398
	1 000	0.38	0.5076	0.42	0.5231
4	250	0.46	0.5105	0.25	0.6491
	500	0.38	0.5990	0.33	0.6565
	750	0.38	0.5368	0.33	0.6204
	1 000	0.15	0.629	0.25	0.6166

Table 5.9 Clutter data from Waldman's direct vision model

5.2.2. Nominal Results

a) Waldman's Model Check

The VISDET model was implemented in MATLAB™ and the results of the model compared to the Waldman's [89] published results. As Waldman's model is independent of the actual scene it makes it possible to do a comparison. The nominal model was called VDnom.m and a listing appears in Appendix B. The search conditions that were given by Waldman were a target of 3mx3m with 8% reflectance against a 10% reflectance background and for ranges varying between 200m and 4000m. The search area was 20 degrees in azimuth and 5 degrees in elevation and lasted 30 seconds. The search was done in full daylight, with light clutter and 12km visibility. Earlier in his paper Waldman described light clutter as having a value of around 0.05. The following assumptions had to be made for the VDnom model - a clutter value of 0.05, a background luminance of 3 420nt (1000 fL) and a sky/background ratio of unity. The results are shown in figure 5.15. The plots in figure 5.15 depict the relative close correlation between the VDnom MATLAB™ model's results and Waldman's published data.

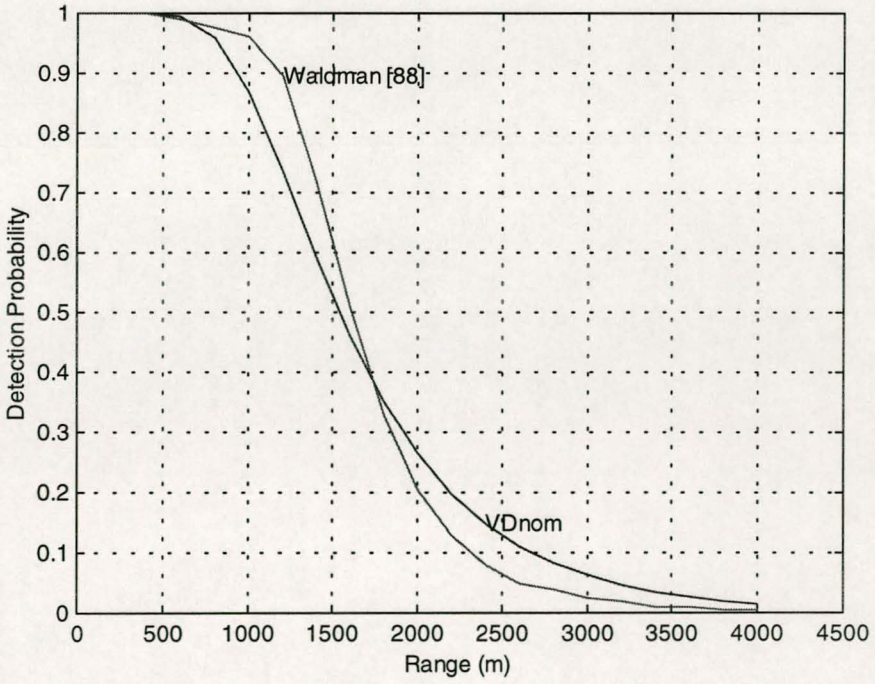


FIGURE 5.15 VISDET NOMINAL ACCEPTANCE RESULTS

b) Nominal ESFID Results for Waldman's Grid Fixation Model

The VISDET model was updated to model the ESFID system but still used the grid model for fixations as developed by Waldman. The main difference between the direct vision model and the ESFID model was the inclusion of the electro-optic display. The light level from the display was limited to 257nt (75fL) and the modulation transfer function of the display was included in the contrast propagation model. The spatial frequency for which the MTF was calculated corresponded to the period defined by Johnson for the target recognition criteria, namely four cycles across the minimum target dimension divided by the range to the target (See table A1). The model was expanded to include both the wide field and the narrow field displays of the ESFID display. The MTF of the target on the display was determined as a function of the display used. If the target was on the narrow field display, the narrow field MTF was used, otherwise the wide field MTF was used. Each of the targets that were used in the ESFID experiment was modelled using the clutter values of table 5.9. The program listing is given in Appendix B under VDESFID1.m.

The acquisition results predicted by the grid model are given in figures 5.18 and 5.19 for the 20% and 30% contrast targets respectively as a function of range. The plots also give the experimental results, together with error bars depicting the one-sigma standard deviation predicted by the binomial distribution.

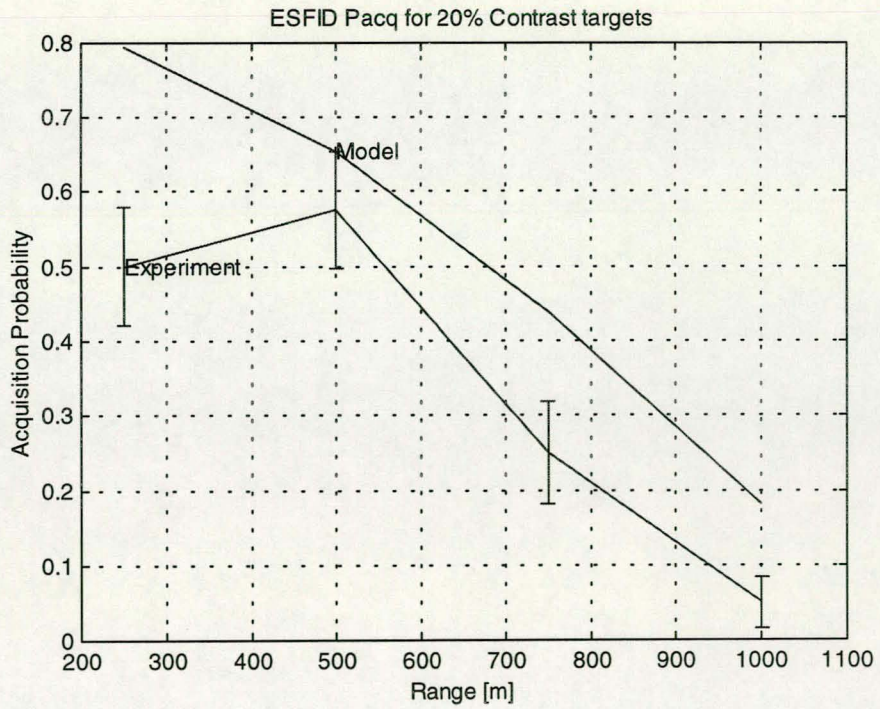


FIGURE 5.16 ESFID1 MODEL PREDICTION WITH GRID FIXATION FOR 20% CONTRAST TARGET

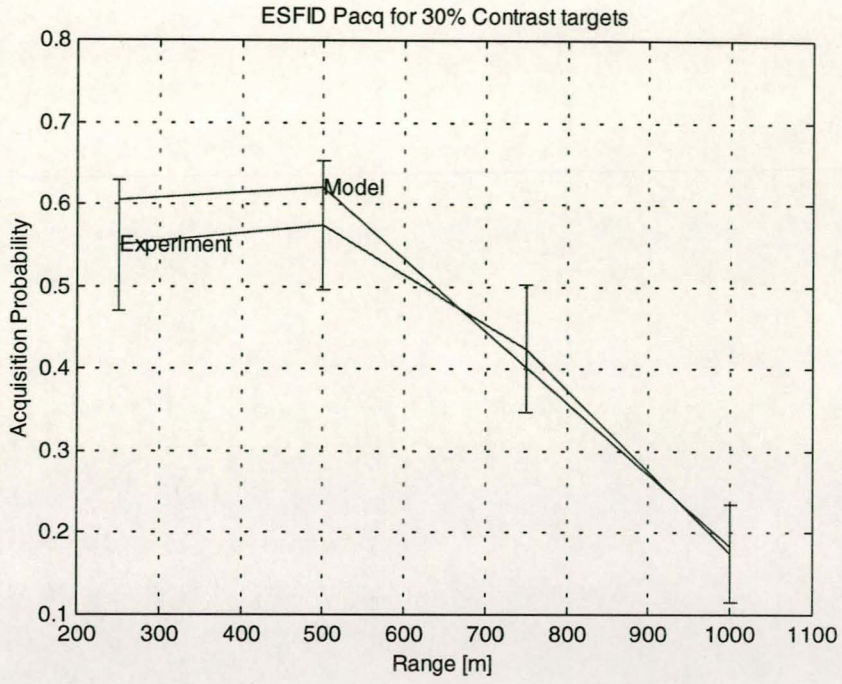


FIGURE 5.17 ESFID1 MODEL PREDICTION WITH GRID FIXATION FOR 30% CONTRAST TARGET

The ESFID grid fixation model over-predicts all the 20% contrast target acquisition probabilities by a relative large margin. The 30% contrast target acquisition probabilities are predicted within the one-sigma experimental error margin.

c) Nominal ESFID Results for Stark's Markov Fixation Model

The fixation model was updated to include the predicted fixations from the Markov model. The single glimpse target detection probability was determined at every fixation, and the cumulative detection probability was determined as a function of the search time. As this is a statistical model, a total of 40 search runs were simulated for each target, and the mean acquisition probability calculated. The cumulative mean acquisition probabilities were plotted in figure5.18 to show the convergence as a function of the number of subjects. The MATLAB™ listing was called VDESFID2.m and it is given in Appendix B.

The results of the Markov fixation model are shown on figures 5.20 and 5.21 for the 20% and 30% contrast targets respectively. Again the experimental results with their one-sigma error bars were plotted for comparative purposes.

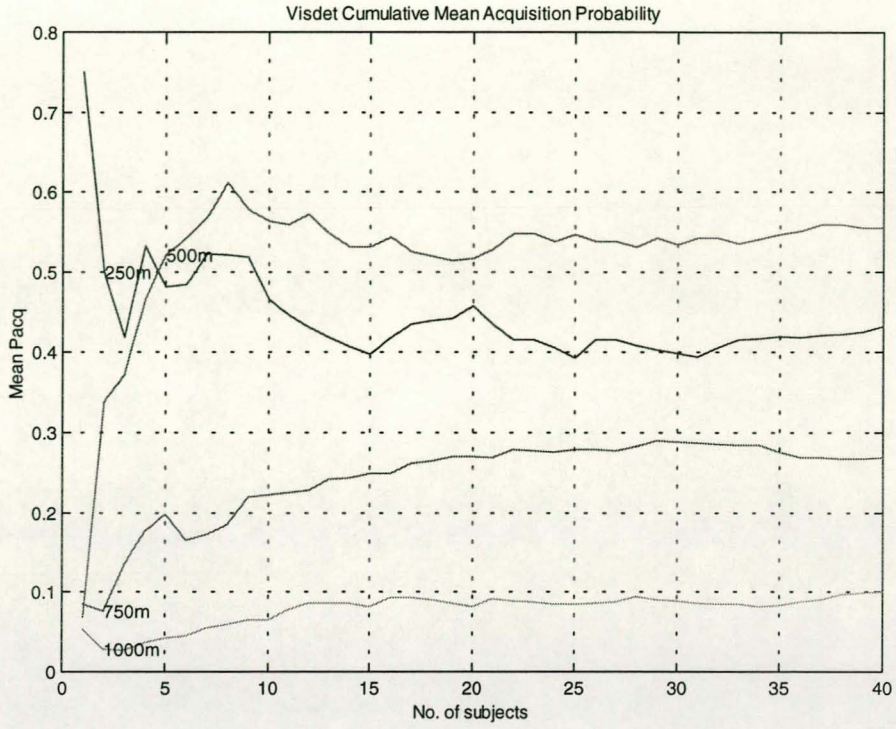


FIGURE 5.18 CUMULATIVE MEAN ACQUISITION PROBABILITY VS. NUMBER OF SUBJECTS

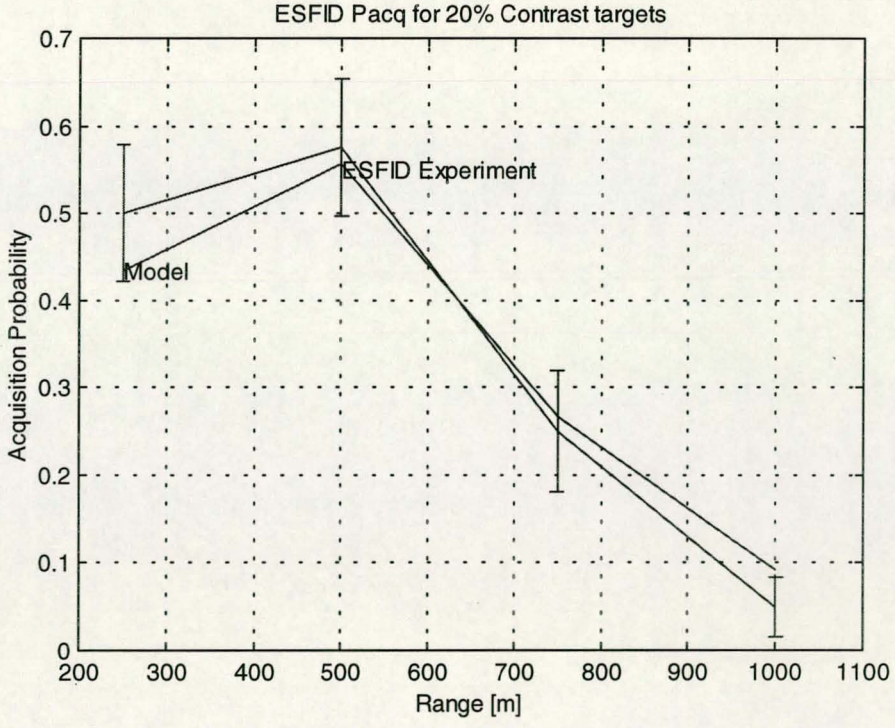


FIGURE 5.19 ESFID2 MODEL PREDICTION WITH GRID FIXATION FOR 20% CONTRAST TARGET

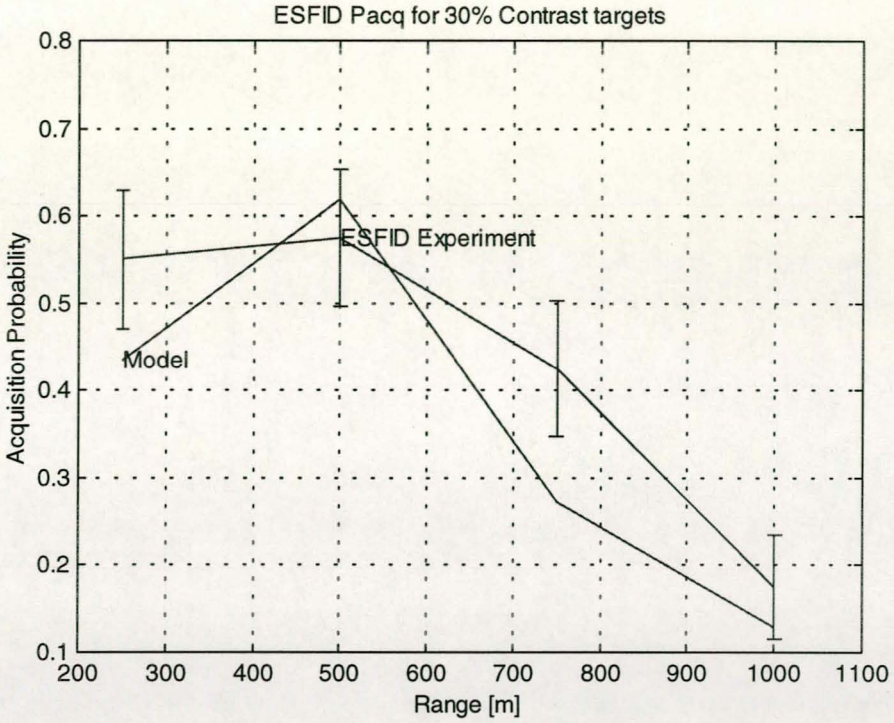


FIGURE 5.20 ESFID2 MODEL PREDICTION WITH GRID FIXATION FOR 30% CONTRAST TARGET

The Markov fixation model gave excellent predictions for the 20% contrast targets. It was only slightly optimistic for the 1000m range target, but the prediction was much better than the grid fixation model. For the 30% contrast targets the correlation was good, but the Markov fixation model under predicts performance at 250m and 750m range.

It was concluded that the Markov fixation model gave a closer match to the actual observed results for a 20% target contrast than Waldman's original grid fixation model. However, Waldman's grid fixation model gave a better match for 30% contrast targets.

5.2.3. Sensitivity Study

The purpose was to determine the sensitivity of the ESFID system to the display design parameters. The performance measure that was used was the mean target acquisition probability as a function of target range. The following major design parameters were identified:

- Field of view
- Resolution (number of display lines on monitor)
- Display Luminance
- Number of resolution cells required for target recognition

a) Field of View

The sensitivity to the field of view was determined using the Markov fixation model. Three narrow fields of view were used - the smallest field of view, of 4.1°x5.47°, corresponds to the limiting eye resolution of the human visual system on a 400-line monitor. The second was the nominal field of view used in the experiments and the third, a field of view of 9.28°x16.5°, corresponds to the high definition TV display described in paragraph 1.4.2 c) of this dissertation. The wide fields of view were scaled along with the narrow fields. The results are shown on table 5.10

Field of View	20% Target Contrast Target Range [m]				30 % Target Contrast Target Range [m]			
	250	500	750	1000	250	500	750	1000
NF 4.1°x5.47° WF 23°x30.7°	0.403	0.59	0.301	0.23	0.421	0.648	0.311	0.347
NF 6.67°x8.9° WF 37.5°x50.0°	0.433	0.556	0.266	0.092	0.434	0.619	0.273	0.129
NF 9.28°x16.5° WF 46.4°x82.5°	0.44	0.488	0.186	0.004	0.437	0.559	0.208	0.006

Table 5.10 Target Acquisition Sensitivity to the Displayed Fields of View

The results showed that for a fixed number of lines on the monitor, the small field of view gave much better target acquisition performance at longer ranges than the nominal field of view, independent of target contrast. At the very short range the wider fields of view performed marginally better than the smaller fields of view. It may be concluded that the overall performance could be increased by using a smaller field of view, but this performance gain may also be ascribed to the effective higher angular resolution that a smaller field of view would give with a fixed number of lines on the monitor. Therefor, the effect of resolution is studied next.

b) Resolution

The resolution of the ESFID experimental displays were determined by the Sony CRT's and was given as 400 lines on the monitor. The sensitivity of target acquisition performance to resolutions that correspond to PAL TV (577 lines) and high definition TV (1080 lines) was determined with the Markov fixation model and the results are shown on table 5.11.

Number of Lines on the Monitor	20% Target Contrast Target Range [m]				30 % Target Contrast Target Range [m]			
	250	500	750	1000	250	500	750	1000
400	0.433	0.556	0.266	0.092	0.434	0.619	0.273	0.129
577	0.436	0.592	0.297	0.234	0.438	0.649	0.306	0.314
1080	0.443	0.614	0.315	0.318	0.446	0.668	0.323	0.423

Table 5.11 Target Acquisition Sensitivity to Monitor Resolution.

The PAL monitor's resolution (577 effective lines over a 6.67° vertical FOV) and the small field of view (400 lines over a 4.1° vertical FOV) of the previous paragraph gave the same angular resolution of 150μrad/pixel (0.52 arcmin/pixel). From table 5.8 and 5.9 it was seen that the thousand-meter target acquisition performance was similar for the PAL resolution and the small FOV conditions. A higher number of lines on the monitor increased the target acquisition performance, especially at long range. The sensitivity at short range was small. The conclusion drawn from these results was that higher angular resolution increased performance.

c) Display Luminance

Display luminance was another design factor in the display design. The luminance capability of display technology is orders of magnitude lower than direct sunlight. Although the threshold contrast of the human eye changes very little between a typical display's luminance and direct sunlight, Waldman modelled the para-foveal contrast threshold of the eye as a function of luminance (See equation A.14 in Appendix A). The nominal ESFID2 model assumed a maximum display luminance of 257nt (75fL). For the sensitivity study this number was doubled in the Markov fixation model. The results are given on table 5.12.

Luminance [nt / fL]	20% Target Contrast Target Range [m]				30 % Target Contrast Target Range [m]			
	250	500	750	1000	250	500	750	1000
257 / 75	0.433	0.556	0.266	0.092	0.434	0.619	0.273	0.129
514 / 150	0.433	0.554	0.265	0.09	0.434	0.617	0.272	0.127

Table 5.12 Target Acquisition Sensitivity to Monitor Luminance

The conclusion was that the luminance level had a negligible effect on target acquisition performance.

d) Number of Resolution Cells Required for Recognition

The number of resolution cells required for recognition (Nrc) was not strictly a design parameter, as it is a function of the human visual system. The literature [31, p 437] did however indicate that the value of Nrc depended on the difficulty of the task. The FLIR92 model also indicated that the two dimensional discrimination task might require fewer resolution cells than Johnson's original single dimensional criteria which was based on the minimum target dimension [31, p436]. The sensitivity of the ESFID display performance was thus also determined for the number of resolution cells, varying between 3 to 5, with 4 being the nominal value.

Number of Resolution Cells on Target	20% Target Contrast Target Range [m]				30 % Target Contrast Target Range [m]			
	250	500	750	1000	250	500	750	1000
3	0.435	0.586	0.29	0.197	0.437	0.645	0.298	0.264
4	0.433	0.556	0.266	0.092	0.434	0.619	0.273	0.129
5	0.431	0.515	0.229	0.018	0.431	0.583	0.244	0.028

Table 5.13 Target Acquisition Sensitivity to Number of Resolution Cells Required for Recognition

The number of resolution cells required for recognition had a significant effect at longer ranges, where the angular size of the target was small. At ranges less than 750 m the effect was relatively small. Comparing the experimental data, it was concluded that the model's assumption of Johnson's criteria of 4 cycles was within the experimental error range.

5.3. GENERALISATION OF RESULTS

5.3.1. Generalisation of the ESFID Results

The ESFID results were generalised for conditions other than those tested in the ESFID experiments to find a design that would confirm the thesis statement. The generalisation of the ESFID results was done with the aid of the VDESFID mathematical models and the sensitivity study of the previous paragraph.

The sensitivity study showed that the angular resolution of the display was important. Due to cost limitations the ESFID display was built around display CRT's that gave only 400 lines resolution. If we postulate high definition TV displays for the ESFID concept, the number of lines on the displays will be 1080. To achieve the 150μrad/pixel (0.52 arcmin/pixel) angular resolution that was shown in the sensitivity study to give the best performance, the narrow field of view would have to be 9.28°x16.5° if a 16:9 aspect ratio is used. The wide field of view would have to be 46.4°x 82.5° to give matching peripheral resolution to the eye, as was shown in paragraph 1.4.2c).

The predicted performance for the high definition display parameters using the Markov fixation model was shown in figure5.21 for the 20% and 30% target contrasts, compared to the direct vision test result. Although the predicted HDTV ESFID performance was close to that of direct vision, it was still below the DVT experiments.

The Markov fixation probability model was based on the fixations measured by the ESFID experiment. The fixations were a function of the displayed scene, and the fixations might have been different if a much higher angular resolution display was used that revealed more detailed information in the scene. For this reason the HDTV parameters were also included in the VISDET1 model, which assumed a grid fixation pattern. The results are shown in figure5.22 for the 20% and 30% target contrasts and compare very closely with the direct vision experimental test results. Target acquisition performance of an ESFID system has been extrapolated using the grid fixation model with a HDTV display, that shows that target acquisition close to that of direct vision could be achieved. It could also be concluded that the grid fixation model had fewer limitations to its generalisation, as it was independent of a specific scene and the fixation points in that scene.

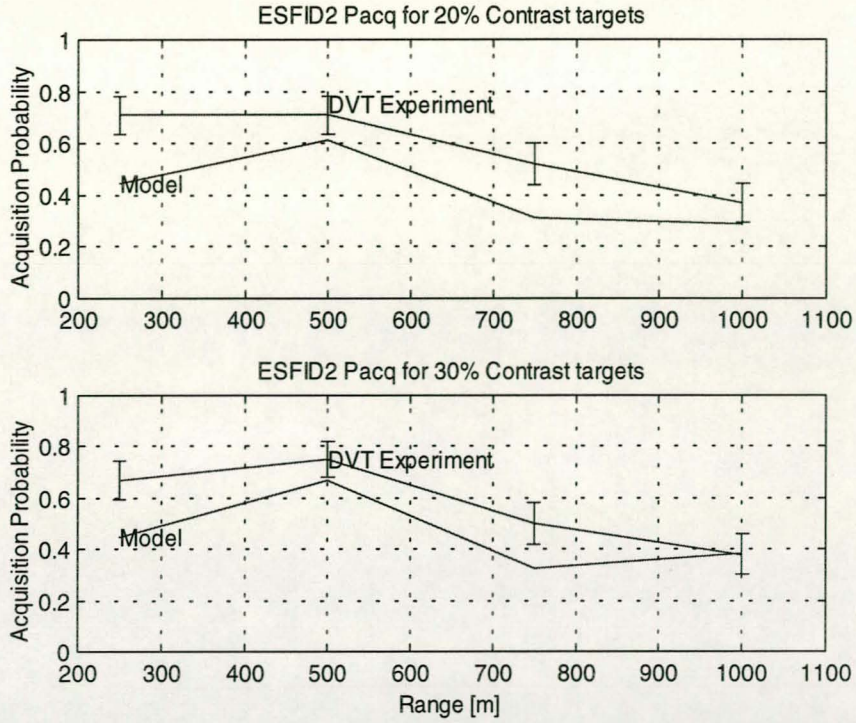


FIGURE 5.21 HIGH DEFINITION TV ESFID PERFORMANCE FROM THE VISDET2 MARKOV FIXATION MODEL

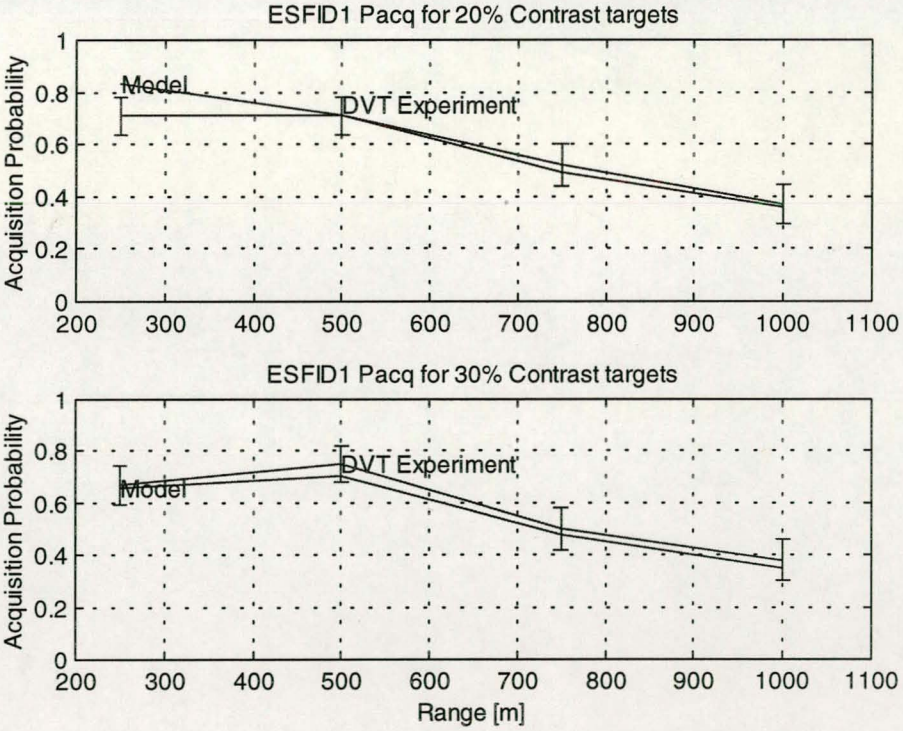


FIGURE 5.22 HIGH DEFINITION TV ESFID PERFORMANCE FROM THE VISDET1 GRID FIXATION MODEL

5.3.2. Identification of where system provides benefits

From the experimental results and the generalisation the following preliminary conclusions can be drawn.

i) The ESFID concept provides superior target acquisition performance to joy-stick controlled displays for all target ranges and contrasts. The mean target acquisition performance of the TDS was 20.6%, compared to 37.8% for the experimental ESFID system, an 80% increase in performance.

ii) The ESFID concept with HDTV displays will provide similar target acquisition performance to direct vision.

iii) The ESFID concept with HDTV displays will provide similar visual resolution to direct vision.

The area where the ESFID concept provides benefits is where remote vision is required with a wide field of view, and with the same resolution as direct vision. Tasks that require high resolution include surveillance, target acquisition, and when combined with accurate remote manipulation, precision teleoperated tasks such as remote surgery, bomb disposal and hazardous operations, construction and repairs. Hazardous environments include military battlefields, nuclear power installations, space, deep-sea operations, mining, and toxic waste disposal plants. An intriguing application of ESFID technology is in scaling man-in-the-loop operations to either very large scale or very small scale. It is through micro manipulation that teleoperated machines could make a contribution to the emerging field of nanotechnology.

The following products give some examples of where the ESFID concept can provide direct benefits:

i) Military operations
- Teleoperated land vehicles such as tanks and reconnaissance vehicles.
- Unmanned combat aircraft, to give the pilot the same visual capabilities as if he was present in the aircraft.
- Unmanned armoured vehicle turrets, which would display a smaller silhouette of the vehicle, while the crew can enjoy the safety of staying inside the armoured shell of the vehicle.
- "Through armour" mast mounted sights on armoured vehicles to provide panoramic observation of the battlefield so that crew-members can get the same visual information inside the vehicle as if they were looking out of the turret.

ii) Remote surveillance: - Security surveillance cameras to give high resolution images of objects of interest such as vehicle number plates, suspects' faces etc., while still covering the whole area of interest

- Airborne observation payloads to give human search performance over a vast search area with day/night capability.

iii) Remote operations: - Teleoperated surgery, where extremely good visual capability is required, together with precision manipulation. Stereopsis will be an added requirement and the concept has to use a display similar to the one in fig. 1.8.

- Bomb disposal, which requires precision vision to identify components.
- Space construction, where peripheral vision is required to orientate parts.
- Deep-sea construction has a similar requirement to space construction.
- Remote mining. Removal of rock after blasting is an extremely hazardous task that causes a large number of fatal accidents in the mining industry every year. Remote controlled machinery will save lives, and the addition of a visual capability similar to direct vision will ensure more natural training and higher performance.
- Nuclear power plant inspection and maintenance. Very small cracks in pipes and other defects have to be found over a large search area. This problem is similar to the target acquisition problem that was studied in this dissertation.
- Scaled operations. Either large scale construction or micro manipulation of biological or nanotechnology specimens require a level of precision that can best be done if the visual senses are fully stimulated with respect to resolution and field of view.

Another application of eye-slaved display technology is in the area of virtual reality, where the computation requirements are dramatically reduced to give a high-resolution scene simulation. The work in this thesis builds on the work of Watson et al. [69], and shows that it is feasible to generate a high-resolution synthetic environment display with only two visual channels.

CHAPTER 6 CONCLUSIONS

6.1. SUMMARY

This dissertation addresses the problem of remote vision. The ideal remote vision system collects all the visual wave front information that the human visual system can sense at a remote site, transmits the information to a remote operator, and reconstructs the information in real time on a 360 degree, all-round display system. The operator is not able to visually discern a difference in being at the remote site or not. Human vision has a peak resolution of between 150-300 μrad (0.5 – 1 arcmin) depending on factors such as the luminance level of the target being viewed. An omni-visual display with a 300μrad (1 arcmin) resolution, will require more than 440 million pixels that has to be updated at an ideal rate of around 50 Hz. The remote vision challenge is to achieve these idealised specifications, or their equivalent. Current and near future technology recorders, communications systems and displays cannot meet the resolution and bandwidth requirements of an ideal remote vision display.

In this thesis a teleoperated sensor design was presented that achieved comparable performance with direct human vision on a specific target acquisition task. The design approximated the above-mentioned ideal information display performance through an eye-slaved foveal inset display (ESFID) design. The design exploited some of the physiological aspects of the human visual system, to approach the idealised display performance. The experimental work was done for monoscopic and monochrome applications, to minimise the number of variables in the experiments.

The conclusions from this thesis were based on a number of assumptions that were tested experimentally and through existing theory. The first assumption was that finite pixel wide-field (e.g. 600 pixels/60°) information and finite pixel narrow-field (e.g. 600 pixels/10°) information could be displayed into separate eyes with the human brain fusing the two pictures into an apparent 6000 pixel display. A cardboard tester provided the experimental proof with direct vision and perfectly aligned channels. The cardboard experiment was followed with an electro-optic, split-eye foveal/peripheral display test that determined that the fusion also occur for finite pixel electro-optic displays. The alignment and magnification tolerances for fusion to occur were also determined. It was shown that it was possible to stimulate all the cone cells of one eye through two video channels, each with a finite pixel resolution display. All test subjects could fuse the two displays successfully, except those who were shown to have strabismus or monoscopic fixation.

The second assumption was that current technology teleoperated sensors gave inferior target acquisition performance compared to direct human vision. This assumption was tested by a set of experiments with a simulated television sighting system and direct human vision on the same natural target and background scenes. The main reasons for the poor performance were ascribed to tunnel vision and the dynamic presentation of the terminal display that reduced the visual acuity of the subjects.

The above ideas led to the thesis hypothesis that by combining head tracking with eye tracking, an eye-slaved foveal inset display system could give similar target acquisition performance to unaided human vision. An actual eye-slaved foveal inset display system was designed and built to test the thesis hypothesis experimentally. It was shown that the narrow field display could be slaved to the human eyeball during target acquisition tasks, without unduly distracting natural vision. The narrow field display was slaved to the eye, during, and for a small time after each saccadic movement of the eye, when the neural activity of the eye was inhibited. Exactly the same target scene as used for the unaided and the television sight simulator was used for comparative target acquisition tests. The experimental results showed that the ESFID concept gave significantly better target acquisition performance than joystick controlled terminal display systems. The experimental ESFID target acquisition results were, on average, below that of direct vision. Cost constraints on the ESFID experimental hardware limited its performance to less than full TV quality and thus the resolution requirements of the ESFID concept, so the final verification of the thesis hypothesis was done through the validated theory.

The empirical model of Waldman et al. [102] was tested, and validated for the teleoperated target acquisition task. It was found that Waldman's model gave an accurate model of eye fixations during search for 30% target contrasts but not for 20% target contrasts. An eye fixation model based on the work of Stark and Ellis [80] was used to extend Waldman's model. Markov chains were used to statistically model fixation sequences, based on the experimental measurement of fixations by the eye-slaved display. Using Stark's fixation model with calibrated data in Waldman's model gave good correlation with all the experimental results. It was discovered that Stark's Markov fixation model had a smaller region of validity, as it was dependent on the actual fixations in the scene, which were a function of the amount of information presented in the display. For the situation where the display information was increased, it was found that Waldman's grid fixation model gave a wider generalisation. The models were used to generalise the experimental results and to predict performance gains for

an eye-slaved display with a high definition TV display. The models predicted that HDTV displays would give similar target acquisition performance to direct vision.

The final result shows that it is possible, by using the ESFID concept with HDTV displays, to build a teleoperated sensor that gives similar target acquisition performance to unaided human vision. The applications where the ESFID system provide benefits over current teleoperator technology has been identified in section 5.3.2. In summary, the ESFID concept can be beneficially used in all applications where remote vision is required with a similar resolution capability and field of view of direct human vision. Virtual reality and simulator displays are also areas where the ESFID concept can provide direct benefits.

6.2. RESEARCH CONTRIBUTION

The research documented in this thesis contributes to the field of teleoperation, helmet mounted display systems, and virtual reality. The research contribution can be allocated to the following three areas:

6.2.1. Split-eye Foveal/Peripheral Displays

The first contribution is the testing of the split-eye foveal/peripheral display concept. The successful implementation of this concept led to the development of a much simpler design for an eye-slaved display than a combined wide and narrow-field display. The work to validate this concept has been presented in [34].

6.2.2. Eye-slaved foveal inset display

The second contribution is the understanding and building of an eye-slaved foveal inset display system and it's testing in a teleoperated target acquisition task. The experiments show that the concept of eye-slaved display works and its performance is close to direct vision. It is also shown by means of a mathematical model, that the teleoperator design can be improved to give the similar target acquisition performance as direct vision. The work on eye-slaved display testing was presented in [107].

6.2.3. Target Acquisition Modelling

The third contribution is the testing of, and expansion of target acquisition modelling, as applied to teleoperated target acquisition systems. The empirical model of Waldman et al. [102] was tested, and validated for the teleoperated target acquisition task. It was found that Waldman's model gave an accurate model of eye fixations during search for 30% target contrasts but not for 20% target contrasts. An eye fixation model based on the work of Stark and Ellis [80] was used to extend Waldman's model. Stark's fixation model gave good correlation with all the experimental ESFID results. It was however discovered that Stark's Markov fixation model was more limited in its generalisation, as it was dependent on the actual fixations in the scene, which were a function of the amount of information presented in the display. For the situation where the display information was increased, it was found that Waldman's model gave a better generalisation.

6.3. CONCLUSIONS

The following conclusions are drawn:

a) It is possible to do target acquisition with an eye-slaved display system of the split-eye ESFID design type. The test subjects accepted the concept of a high-resolution display that followed their eye fixations, as well as a split-eye display with the wide field in one eye and the narrow field in the other eye.

b) The ESFID concept provides superior target acquisition performance to joystick-controlled displays for all target ranges and contrasts. The mean target acquisition performance of the TDS was 20.6%, compared to 37.8% for the experimental ESFID system, an 80% increase in performance.

c) The ESFID concept with HDTV displays will provide similar target acquisition performance to direct vision.

d) The ESFID concept with HDTV displays will provide similar visual resolution to direct vision.

e) It is especially at short range that the ESFID shows big performance gains over current teleoperated sight displays, which makes it ideal for through-armour vision applications.

f) Operators suffering from strabismus or monoscopic fixation cannot use this split-eye eye-slaved foveal inset display system.

g) It was found that Stark's Markov fixation model was more limited in its generalisation than Waldman's grid fixation model. This was especially true for higher display resolution modelling where an increase in the information displayed to the operator could alter his fixation pattern.

6.4. RECOMMENDATIONS

The following recommendations have been made for further work on the eye-slaved display

a) The display should be updated to a lightweight, ergonomically friendly design that leaves the operator's hands free. Such a display will encourage the operators to follow more natural search behaviour in a teleoperated target acquisition system.

b) The effect of colour should be included. Although it should not have a major effect on target acquisition performance, it was a major drawback for telepresence. None of the operators in the system experienced telepresence in the design that was tested. Their comments all indicated that the lack of colour reduced the realism of the presented display. The effect of a sense of telepresence on target acquisition performance should be studied further.

c) The resolution should be upgraded to at least 800 by 600 over the same fields of view, to give close to true human visual resolution. This can be followed with high definition TV resolutions of 1080 by 1920 pixels, with a wider narrow field of view, to cover the full capability of the human visual system.

d) Investigate means to add the narrow field directly to the wide field channel, similar to the concept in figure 1.8, permitting stereo use.

REFERENCES

1. S.Tachi, K. Tanie, K. Komoriya and M. Kaneko (Mechanical Engineering Laboratory, MITI, Japan), "Tele-existence (I): Design and evaluation of a visual display with sensation of presence", Proceedings of RoManSy '84: The fifth CISM-IFToMM Symposium, MIT Press Cambridge, 1984, pp. 245-254.
2. B. A. Wandell, "*Foundations of Vision*", Sinauer Associates, Massachusetts, USA, 1995, p. 49.
3. H.G. Graham, editor, "*Vision and Visual Perception*", John Wiley & Sons, Inc., New York, 1965, p. 253.
4. Ibid., p. 352.
5. W. Barfield, C. Hendrix, O. Bjorneseth, K.A. Kaczmarek and W. Lotens, "Comparison of Human Sensor Capabilities with Technical Specifications of Virtual Environment Equipment", *Presence*, Vol. 4, Fall 1995, pp. 329-356.
6. T. B. Sheridan, "Defining our terms", *Presence,* Vol. 1, No. 2, Spring 1992, pp. 272-274.
7. C.M. Witkowski, A.H. Bond and M. Burton, "The Design of Sensors for a mobile Teleoperator Robot", *Digital Systems For Industrial Automation*, Vol. 2, No. 1, 1983, pp. 85-111.
8. A.K. Bejczy and T.L. Brooks, "Advanced Control techniques for teleoperation in Earth Orbit", Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, Circa 1981.
9. K.E. Chandler, "Underwater video imaging and Remotely Operated Vehicles", *Unmanned Systems*, Spring 1991, pp. 35-38.
10. T.A. Nobbe, "Robots go where no man has gone before", *Machine Design*, January 25, 1990, pp. 52-56.
11. F.P. Rucky, "Remotely controlled Robotic Vehicle For Fire Fighting", *IEEE International Conference on Intelligent Control*, 1987, S.L., pp. 239-245.
12. M. Hewish and G. Turbe, "Europe and US pursue battlefield robotics", *International Defence Review*, No. 1, 1991, pp.40-46.
13. F. Tendick, R.W. Jennings, G. Thorp and L. Stark, "Perception and Manipulation Problems in Endoscopic Surgery", *Computer-Integrated Surgery*, edited R. Taylor et al., The MIT Press, Cambridge, Massachusetts, 1996.
14. A. D. Broadhurst, "Television and Transmission", Dept. of Electrical Engineering, University of Natal.
15. G. Niemeyer and J.E. Slotine, "Stable Adaptive Teleoperation", *IEEE Journal of Oceanic Engineering*,

Vol. 16, No. 1, January 1991, p. 152.

16. G. Green,"Desert Storm, Unmanned air vehicles in the aftermath", *Unmanned Systems*, Summer 1991,p7.

17. R. Byrne, and P.R. Klarer, "Military Robotics Technologies", *Unmanned Systems*, Spring 1992, p. 45.

18. D.J. Busse and K.G. King, "Forward Observation Remote Target Acquisition System (FORTAS)", *Unmanned Systems*, Fall 1992, pp. 14-19.

19. D.P. Miller, "Evaluation of Vision Systems for Teleoperated Land Vehicles", *IEEE Control Systems Magazine*, June 1988, pp. 37-41.

20. RCA Electro-optic Handbook, *RCA Corporation*, 1974, pp.109-125.

21. H.H. Bailey, "Target Acquisition Through Visual Recognition: An Early Model", ADA030699, The Rand Corporation, September 1972.

22. A.V. Barber, "Visual Mechanisms and Predictors of Far Field Visual Task Performance", *Human Factors*, 1990, Vol. 32, No. 2, pp. 217-233.

23. Y. Yeh and L.D. Silverstein, "Limits of Fusion and Depth Judgement in Stereoscopic Colour Displays", *Human Factors*, 1990, Vol. 32, No. 1, pp. 45-60.

24. P. Fiorini, A.K. Bejczy, and P.S. Schenker, "Integrated Interface for Advanced Teleoperation", *IEEE Control Systems*, October 1993, pp. 15-20.

25. W.S. Kim, A Liu, K. Matsunaga and L. Stark, "A Helmet Mounted Display For Telerobotics", COMPCON Spring 1988, 33rd *IEEE Computer Society International Conference*, San Francisco, California, USA, 29 Feb-3March 1988, pp. 543-547.

26. D.P. Miller, "Evaluation of Vision Systems for Teleoperated Land Vehicles", *1987 International Conference on Systems, Man and Cybernetics*, Alexandria, Virginia, October 20-23, 1987.

27. Kenneth N. Ogle, "Vision", *Electrical Engineers Handbook* - Electric Communication and Electronics, John Wiley and Sons, 1950, pp. 14-25 - 14-49.

28. D. Noton and L. Stark, "Scanpaths in eye movements during pattern perception", *Science*, 1971, No. 171, pp. 308-311.

29. K.R. Boff and J.E. Lincoln (Ed.), "Engineering Data Compendium, Human perception and performance", Vol. 1., H.G. Armstrong Aerospace Medical Research Laboratory, Wright Patterson Air Force Base, Ohio, 1988, section 1.602, p199.

30. O.H. Schade Sr., "A new System of Measuring and Specifying Image Definition", National Bureau of Standards Circular 526, 231-258, 1954.

31. Gerald C. Holst, "*Electro-Optical Imaging System Performance*", JCD Publishing, Winter Park, FL and SPIE Optical Engineering Press, Bellingham WA, USA, 1995, p. 148.

32. G. Westheimer and S.P. McKee, "Stereoscopic acuity with defocused and spatially filtered retinal images", *Journal of the Optical Society of America*, Vol. 70, 1980, pp. 772-778.

33. B. Bhatt, D. Birks and D. Hermreck, "Digital television: making it work", *IEEE Spectrum*, Oct 1997, p20.

34. G.T. Viljoen, "A Split Foveal/Peripheral Display for Monocular Teleoperator Control Systems", *Proceedings of the IEEE Small Satellite and Control Conference*, Stellenbosch, South Africa, Oct. 1994.

35. F.C. Volkmann, A.M. Schick, and L.A. Riggs, "Time course of visual inhibition during voluntary saccades", *Journal of the Optical Society of America*, 1968, No.58, pp. 562-569.

36. T.B. Sheridan, "Telerobotics", *Automatica*, Vol. 25, No. 4, pp. 487-507, 1989.

37. Edwin G. Johnsen and William R. Corliss, *Human Factors Application in Teleoperator Design and Operation*, John Wiley & Sons Inc., 1971, p7.

38. Jean Vertut and Philipe Coiffet, *Teleoperation and Robotics: Vol. 3A Evolution and Developmen*", Prentice Hall, Englewood Cliffs, New Jersey 07632, 1986.

39. Jean Vertut and Philipe Coiffet, *Teleoperation and Robotics: Vol. 3B Applications and Technology*, Prentice Hall, Englewood Cliffs, New Jersey 07632, 1986.

40. C.P.Comeau and J.S. Bryan, "*Headsight television systems provides remote surveillance*", Electronics, November 1961, Vol. 34, pp. 86-90.

41. I.E. Sutherland, "The ultimate display", International Federation of Information Processing Congress, 1965, Vol. 2, pp. 506-508.

42. B.A. Chouet and L.R. Young, "Tracking with Head Position Using an Electro-optical Monitor", *IEEE transactions on Systems, Man and Cybernetics*, VOL.SMC-4, No. 3, March 1974, pp.192-204.

43. D.K. Shirachi, D.L. Monk and J.H. Black,, "Head Rotational Spectral Characteristics During Two-dimensional Smooth Pursuit Tasks", *IEEE Transactions on Systems, Man and Cybernetics*, VOL.SMC-8, No. 9, September 1978, pp. 715-724.

44. F. Hamit, *Virtual Reality and the Exploration of Cyberspace*, SAMS publishing, Carmel, Indiana, USA, 1993, pp. 61-65.

45. P. Wetzel and M. Eisenmann, "*Engineering and human vision considerations in the design of the FOHMD*", Proceedings of the Royal Aeronautical Society, Advances in Flight Simulation, London, 1986.

46. P.W. Goode "Eye Directed View", Automation Technology Branch, Langley Research Centre, NASA,

Circa 1984.

47. S. Tachi, H. Arai, I. Morimoto and G. Seet, "Feasibility experiments on a mobile tele-existence system", *Proceedings of the International Symposium and Exposition of Robots*, 19th ISIR by the International Federation of Robotics, Sydney, NSW, Australia, 6-10 Nov. 1988, pp.625-636.

48. William R. Uttal, "Teleoperators", *Scientific American*, December 1989, pp. 74-78.

49. John R. McDonnel, Manuel R. Solorzano, Stephen W. Martin and Alan Y. Umeda (NOSC - Naval Ocean Systems Centre), "A head coupled sensor platform for teleoperated ground vehicles", *Unmanned systems*, Fall 1990, pp. 33-38.

50. S.R.Ellis, "*Review of Virtual Reality by Howard Rheingold*", Ergonomics, 1993, 36, 6, pp. 743-46

51. F. Brooks Jr., "Grasping reality through illusion – Interactive graphics serving science", Proceedings of SIGCHI'88, May 15-19 1988, Washington D.C., pp. 1-12.

52. F. Quek, R. Jain and B. Mitchell, "Tele-perception", *SPIE*, Vol. 1006, Space Station Automation IV, 1988, pp.152-162.

53. H. Yamaguchi, A. Tomono, Y. Kobayashi, "Proposal for a Large Visual Field Display Employing Eye Movement Tracking", Proceedings of SPIE Vol. 1194, Optics, Illumination, and Image Sensing for Machine Vision IV, Philadelphia, PA, USA, 8-10 November1989, pp. 13-20.

54. R. Finkelstein, "Telepresence and Virtual Reality", *Unmanned systems*, Summer 1989, pp.7-8.

55. T.B. Sheridan, "Musings on telepresence and virtual presence", *Presence*, Vol.1, No.1, 1991, pp.120-125.

56. R. M. Held and N.I. Durlach, "Telepresence", *Presence*, Vol. 1, No. 1, 1991, pp.109-112.

57. S.R. Ellis, M.K. Kaiser, A.J.Grunwald, (editors), "Pictorial communications in virtual and real environments", Taylor and Francis, London, 1991, pp. 232-246

58. J. Hall, "Unmanned Ground Vehicles - An Update", *Unmanned Systems*, Spring 1992, pp. 16-19.

59. D. Rasmussen, "A Telepresence Interface For Precision Telerobot Control", NASA Ames research centre, Moffet Field, CA, circa 1993.

60. A. Liu, G. Tharp, L. French, S. Lai and L. Stark, "Some of What One Needs to Know About Using Head-Mounted Displays to Improve Teleoperator Performance", *IEEE Transactions on Robotics and Automation*, Vol. 9, No. 5, October 1993, pp. 638-648.

61. K. Meyer, H.L. Applewhite, and A. Biocca, "A Survey of Position Trackers", *Presence*, Vol. 1, No. 2, Spring 1992, pp. 173-200.

62. M. McKenna and D. Zeltzer, "Three Dimensional Display Systems for Virtual Environments", *Presence*,

Vol. 1, No. 4, Fall 1992, pp. 421-458.

63. J. Liang, C. Shaw and M. Green, "On temporal-spatial realism in the virtual reality environment", *Proceedings of the Fourth Annual Symposium on User Interface Software and Technology*, Hilton Head, South Carolina [UIST}, November 1991, pp. 19-25.

64. J. Chung, M. Harris, F. Brooks, et al., "Exploring virtual worlds with head-mounted displays, non-holographic 3-dimensional display technologies", *SPIE Proceedings, Non-holographic 3-Dimensional Display Technologies*, Los Angeles, CA, 15-20 January 1989, p.1083.

65. R. Held and N. Durlach, "Telepresence, time delay and adaptation", S. Ellis (ed.), *Representation in pictorial and virtual environments*, Taylor and Francis, London, 1991.

66. S.R. Ellis, N.S. Dorighi, B.M. Menges, B.D. Adelstein and R.H. Jacoby, *"In search of equivalence classes in subjective scales of reality*", Proceeding of HCI'97 International, 1997, San Francisco pp. 873-876.

67. S.R. Ellis, F. Breant, B.M. Menges, B.D. Adelstein and R.H. Jacoby, *"Operator interaction with virtual objects: effects of system latency*", Proceeding of HCI'97 International, 1997, San Francisco, pp. 973-976.

68. S.R. Ellis, M.J. Young, B.D. Adelstein and S.M. Ehrlich, *"Discrimination of changes of latency during voluntary hand movement of virtual objects* ", Proceeding of Human Factors and Ergonomics Society, September 27 – October 2, 1999, Houston Texas.

69. B. Watson, N. Walker and M. Reddy, "An Evaluation of Level of Detail Degradation in Head-Mounted Display Peripheries", *Presence*, Vol. 6, No. 6, December 1997, pp. 630-637.

70. D.H. Hubel, *Eye, Brain and Vision*, Scientific American Library, New York, 1988.

71. S. Shah and M.D.Levine, "Visual Information Processing in Primate Cone Pathways - Part I: a model", *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics*, Vol. 26, No. 2, April 1996, pp. 259-274.

72. A.Baveja, C.G.Drury, M.H.Karwan and D.M.Malon, "Derivation and Test of an Optimum Overlapping-Lobes Model of Visual Search", *IEEE Transactions on Systems, Man and Cybernetics-Part A:Systems and Humans,* Vol. 26, No. 1, January 1996, pp. 161-173.

73. A.Ford, C.T.White, and M.Lichtenstein, "Analysis of Eye Movements during Free Search", *Journal of the Optical Society of America*, Volume 49, Number 3, March 1959, pp287-292.

74. A.T. Bahill, D. Adler, L. Stark, *"Most naturally occurring human saccades have magnitudes of 15 degrees or less*", Investigative Ophthalmology, 1975, Vol. 14, pp. 468-469.

75. S.W.Smith, "Visual Search Time and Peripheral Discrimination", *Journal of the Optical Society of America*, Vol. 51, 1961, p 1462.

76. D. Noton and L.W.Stark, "Eye Movements and visual perception", *Scientific American*, Vol. 221, no. 6, June 1971, pp.34-43.

77. M.V. Srinivasan, M.A.L. Thathachar and B.L. Deekshatulu, "A Probabilistic Hypothesis for the Prediction of Visual Fixations", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-5, No. 4, July 1975, pp. 431-437.

78. J.C. Leachtenauer, "Peripheral Acuity and Photointerpretation Performance", *Human Factors*, Vol. 20., No. 5, October 1978, pp 537-551.

79. E.D. Megaw and J. Richardson, "Target Uncertainty and Visual Scanning Strategies", *Human Factors*, Vol. 21., No. 3, June 1979, pp 303-315.

80. L.Stark and S.R. Ellis, "Scanpaths Revisited: Cognitive Models Direct Active Looking", *Eye Movements: Cognition and Visual Perception*, edited by D. Fisher. R. Monty, J.Senders, Lawrence Erlbaum Associates Publishers, Hillsdale, New Jersey 1981, Part IV.1, pp193-226.

81. J.W. Senders "*A reanalysis of the pilot eye-movement data*", IEEE Transactions on human factors in electronics, 1966, HFES-7, pp103-106.

82. S.R. Ellis and L. Stark, "*Statistical dependency in visual scanning*", Human factors, 1986, Vol. 28, pp. 421-438.

83. S.R. Ellis and L. Stark, "*Eye movements while viewing Necker cubes*", Perception, 1978, Vol. 7, pp575-581.

84. K. Kraiss and A. Knäeuper, "Using Visual Lobe Area Measurements to Predict Visual Search Performance", *Human Factors*, December 1982, Vol. 24, No. 6, pp673-682

85. L.J. Williams, "Cognitive Load and the Functional Field of View", *Human Factors*, Vol. 24, No. 6, December 1982, pp683-692

86. A.J.Courtney, "A Search Task to Assess Visual Lobe Size", *Human Factors*, Vol. 26, No. 3, June 1984, pp289-298

87. A.J.Courtney and C.H. Shou. "Simple measures of visual-lobe size and search performance", *Ergonomics*, Vol. 28, No. 9, September 1985, pp1319-1331

88. A.P.Golding, "Search Performance and Peripheral Visual Acuity", Pretoria: Human Sciences Research Council (Report: National Institute for Personnel Research; PERS-409, ISBN 0796904049), 1986.

89. S.S. Hacisalihzade, L.W. Stark and J.S. Allen, "Visual Perception and Sequences of Eye Movement Fixations: A Stochastic Modelling Approach", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 22, No. 3, May/June 1992, pp. 474-481.

90. G.K.Hung, J. Wilder, F.L.Weiss and R.K.Curry, "Random and direct path eye movements during target search", *Medical Science Research*, Vol. 21, 1993, pp389-391

91. H.R. Blackwell, "Contrast Thresholds of the Human Eye", *Journal of the Optical Society of America*, Vol. 36, No. 11, November 1946, pp. 624-643.

92. A.B. Kristofferson, H.R. Blackwell and S.W. Smith, "Visual Detection as influenced by target form", *Proceeding of a Armed Forces - NRC Committee on Vision Symposium on Form Discrimination as Related to Military Problems*, Ed. J.W. Wulfeck and J.H. Taylor, Tufts University, Medford, Massachusetts, April 4-5, 1957, pp. 109-127.

93. E.S.Krendel and J.Wodinsky, "Search in an Unstructured Visual Field", *Journal of the Optical Society of America*, Vol. 50, No. 6, June 1960, pp. 562-568.

94. J.R.Bloomfield, "Visual Search in Complex Fields: Size Differences Between Target Disc and Surrounding Discs", *Human Factors*, Vol. 14, No. 2, April 1972, pp.139-148.

95. H.H.Bailey, "Target Detection Through Visual Recognition – a Quantitative Model", RM-6158-PR, *The Rand Corporation*, February 1970.

96. E.K.Seyb, "Influence of search mode on visual detection of airborne target", *SHAPE technical Memo STC TM-458*, 1975.

97. H.W. Wessely, "Limiting Factors in Tactical Target Acquisition", *Rand Report no P-5942V*, March 1978.

98. L.Olzak, "Detection of Multiple Vehicle Targets in Realistic Terrain", Hughes Aircraft Corporation for *Naval Weapon Centre*, China Lake, CA 93555, Report no. NWC TP6861, July 1978.

99. B.O.Koopman, "Search and Screening: General Principles with Historical Applications", *Pergamon Press*, New York, 1980.

100. B.O.Koopman, "An Empirical Formula For Visual Search", *Operations Research*, Vol.33, No. 3, May-June 1986.

101. K.Wootton, G.Waldman, G.Hobson, and K.Leutkemeyer, "A Normalized Clutter Measure for Images", *Computer Vision, Graphics, and image processing*, 1988, No. 42, 137-156.

102. G.Waldman, K.Wootton and G.Hobson, "Visual Detection with Search: An Empirical Model", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 21, No. 3, May/June 1991, pp. 596-606.

103. J.M.Enoch, "Effect of the Size of a Complex Display upon Visual Search", *Journal of the Optical Society of America*, Vol. 49, No. 3, March 1959, pp. 208-286.

104. T.J.Doll, S.MacWhorter, D.E.Schmieder, A.D. Sheffer, and W.R.Owens, "Simulation of human vision for visual/electro-optical (VISEO) detection analysis", Proceedings of the sixth annual Ground Target Modeling and Validation Conference, Keenaw Research Centre, Houghton, MI, 1995.

105. Lawrence Stark, et. al., "Telerobotics: Problems and Research Needs", *IEEE Transactions on aerospace and Electronic Systems*, Vol.24, No.5 September 1988, pp. 542-551.

106. L.J.Hettinger and G.E.Riccio, "Visually Induced Motion Sickness in Virtual Environments", *Presence*, Vol. 1, No. 3, Summer 1992, pp. 306-310.

107. G.T.Viljoen, "Comparative study of target acquisition performance between an eye-slaved foveal inset display system and unaided human vision", *Proceedings of the SPIE Aerosense '98 Helmet and Head-Mounted Displays III conference in Orlando,* 13-17 April 1998, Paper 3362-06.

108. H.R. Blackwell and J.H. Taylor, "Survey of laboratory studies of visual detection", *Paper presented at NATO seminar on Detection, Recognition, and Identification of Line of sight targets*, The Hague, Netherlands, 1969.

109. E.R. Hooton and K. Munson (Editors), Jane's Battlefield Surveillance Systems, Sixth Edition 1994-95, Jane's Information Group Limited, Surrey, UK, pp.102-118.

110. B. Nortje, "A System for Eye-directed Control in a Split Foveal-Peripheral Display", Masters Thesis, University of Pretoria, 1999.

111. C.T. Scialfa, P.M.Garvey, K.W. Gish, L.M. Deering, H.W. Leibowitz, C.C. Goebel, "Relationships among Measures of Static and Dynamic Visual Sensitivity", Human Factors, Volume 30, No. 6, June 1988, pp.677-687.

112. J.W. Miller, "Study of visual acuity during the ocular pursuit of moving objects", *Journal of the Optical Society of America*, Vol. 48, 1958, pp. 803-808.

113. R.E.Walpole and R.H. Myers, "*Probability and Statistics for Engineers and Scientists*", Macmillan Publishing Company, New York, USA, 1985, pp. 290-294.

APPENDIX A

ESFID MATHEMATICAL MODEL DESCRIPTION

A.1 INTRODUCTION

A mathematical model will be used to generalise the experimental results. As the experiments can only be done with limited performance hardware, the mathematical model will be used to extrapolate the results, while staying inside the model's validity region. . The mathematical model follows from the literature. The VISDET empirical model developed by Waldman et al [102] was used. This model was implemented in Matlab and is given in Appendix B, together with its validation results. The model consists of a number of sub-models that will be described in detail in the rest of this appendix. Waldman's fixation model uses a rectangular array of equal probability fixation points, which is shown in the literature to be an unrealistic assumption [76]. The author updated Waldman's model with a more representative fixation model, based on the work of Stark and Ellis [80], which uses a Markov model to model the order of visual fixations.

A.2 FIELD MODEL

The field model predicts the target acquisition probability in the field based on a number of sub-models, which will be described later in this appendix. The field model is described in the block diagram in figure A.1. The field model starts with the scene sub-model of the target against the background that accepts inputs of ambient illumination, background and target reflectance, and the target size. The output of the scene sub-model is the inherent contrast and the scene luminance at zero range. The atmospheric model calculates the received contrast and scene luminance at the system entrance pupil. The field model then splits into three options, one for direct vision, one for an optical system, and one for an electro-optic system. In the case of direct vision, the contrast from the atmospheric model is transmitted straight through to the human sub-model. In the other two cases the atmospheric contrast is reduced with the modulation transfer function of the system at the spatial frequency $1/\lambda d$ where λ corresponds to the Johnston recognition criteria of the target [31]. The human sub-model determines the human factor as a function of the amount of clutter in the scene. The fixation model determines the sequence of fixations, which can either be a grid-like pattern, or from a Markov model of the eye's fixations [80]. The detection model first determines the single shot probability of detecting the target for each fixation. The detection probabilities for all the fixations are summed statistically for the full allowable search period to get the

detection probability. Finally the recognition probability is determined, based on Johnston's criteria. The final acquisition probability is the product of the detection and recognition probability.

The VISDET field model is summarised in block diagram form in figure A.1.

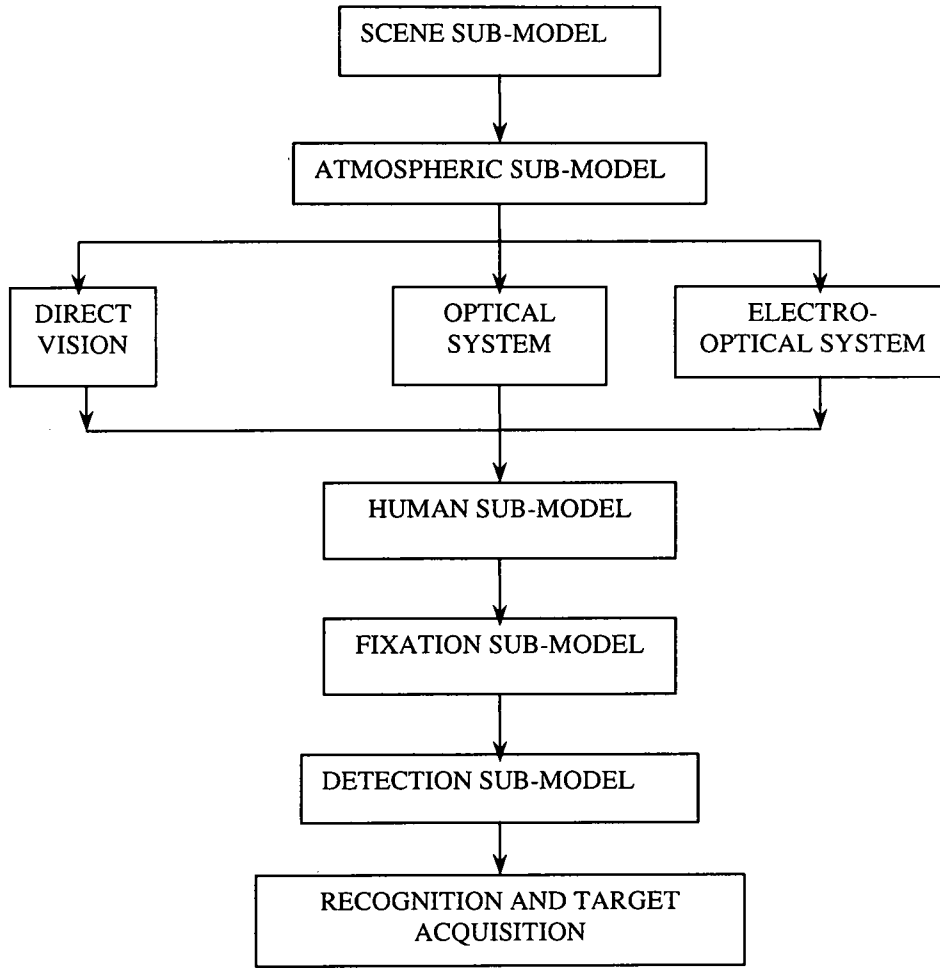


Figure A.1 Block diagram of VISDET target acquisition model

A.3 SCENE SUB-MODEL

The background model is a monochrome digitised image file with 64 grey levels. The target model is a three-dimensional polygon model of a T72 tank that is superimposed on top of the background model. The target lighting model consists of uniformly lit polygons and Lambertian reflecting surfaces. The input to the target model is ambient light level and reflectance of each polygon. The target is overlaid on the background and subsampled with a second order anti-aliasing filter. The target and background model is described in more detail in paragraph 3.4.1.

The output of the scene sub-model is the inherent contrast and scene luminance at zero range. The inherent contrast is determined as follows: A mask is created of the target silhouette by drawing the target in a single colour only, to determine which pixels belong to the target, and which pixels to the background. The size of the target, as measured by the minimum and maximum in the x and y directions, is also determined. Now an area with twice the area of the target (that is 1.41 times the target dimension in x and y) is selected on the background centred on the position where the target is to be drawn. The average background brightness, B_B, is then determined, excluding the area covered by the target, using the target mask. The average target brightness is determined by drawing the target in its proper shades of grey, but on a distinct background (that is, with a colour that was not used for the target). The average target brightness, B_T, is then determined by averaging the brightness over the picture, but excluding the background. The inherent contrast is then determined as follows:

$$C_0 = \frac{|B_B - B_T|}{B_B} \tag{A.1}$$

The scene luminance at zero range is measured during the recording of the background scene. See paragraph 3.4.1 for more details.

A.4 ATMOSPHERIC SUB-MODEL

The atmospheric model uses range to the target, visibility, and precipitation and calculates received contrast and scene luminance at the system entrance pupil. The atmospheric model uses the following simple algorithms [102] that gives the aerosol scattering transmittance as:

$$\tau = \exp\left(-3.91\frac{R}{V}\right) \tag{A.2}$$

where R is the range to the target and V is the visibility range. The model for this dissertation does not include any effect for precipitation such as rain or snow, as all the scenes are clear weather scenes. The propagation of contrast C_R, and scene (background) luminance L_{RB} is governed by

$$C_R = C_0\left[1 - S\left(1 - \tau^{-1}\right)\right]^{-1} \tag{A.3}$$

and

$$L_{RB} = \tau L_{OB}\frac{C_0}{C_R} \tag{A.4}$$

respectively, where C_0 is the inherent contrast, S is the sky to ground brightness ratio, and L_{OB} the background luminance at zero range.

A.5 VIEWING SYSTEM MODEL

The field model splits into three possible sub-models, one for direct view, one for optical (telescopic) view and one for electronic imaging. For this thesis only the direct view and the electronic imaging models are used.

- If the direct view option is used, the received contrast and luminance variables becomes the displayed values, and the human sub-model is directly used with the display angular size set to the whole search sector.
- In the case of an optical sight, the displayed luminance becomes reduced by the transmittance of the optics, the displayed contrast is reduced by the system modulation transfer function (MTF) at the spatial frequency that corresponds to Johnson's criteria for target recognition[31] as shown on table A1, and the displayed target angular size is increased by the magnification. Display angular size is the optical system field of view multiplied by the magnification. The average glimpse probability is determined by averaging over the fixations in a magnified search sector, but omitting those that could not be displayed out of the field of view.
- The chief characterisation of an electro-optical imaging system is the input limiting resolution versus the faceplate illumination curve. The optical input allows calculation of the faceplate illumination that leads to a limiting resolution. A Gaussian MTF curve is internally constructed with this limiting resolution as the 3% point. Then the target contrast is reduced by the MTF at the spatial frequency that corresponds to Johnson's criteria for target recognition[31] as shown on table A1. To get an actual displayed luminance, contrast, and angular size, the maximum display luminance, reproducible shades of grey and operating distance is required. In a manner analogous to the telescopic sight option, the whole search sector must be searched but only a magnified portion is displayed at any time.

In this dissertation, the direct vision model will be used, together with two versions of the electronic imaging systems, corresponding to a terminal display system, and to an eye-slaved display system.

Task	Line Resolution per Target Minimum Dimension
Detection	1.0 ± 0.25 line pairs
Orientation	1.4 ± 0.35 line pairs
Recognition	4.0 ± 0.8 line pairs
Identification	6.4 ± 1.5 line pairs

Table A1 Johnstone's Line Resolution Requirements for a single military target [31]

A.6 HUMAN SUB-MODEL

The human is modelled by the human factor, F_H in equation (A.6). Waldman found that an empirical correction is required in the form of an exponential function which has the desirable properties of giving exactly unity at zero clutter, and never reaching exactly zero at any clutter value.

$$F_H = \exp\left(-5.46 C_L^{2.37}\right) \tag{A.6}$$

In addition to the human factor, human vision has a finite response time before it reacts to stimulus. VISDET models the response time also as a function of clutter:

$$t_r = 0.106 \exp\left(3.57 C_L^{1.8}\right) \tag{A.7}$$

The response times are fast, in the order of 0.1 to 0.2s and compared to the search times of 10-30s that was used in this dissertation, the response time model has a negligible effect.

A.7 FIXATION TIMES AND SEPARATIONS SUB-MODEL

Equation 3.1 determines the specific one-glimpse probability for a specific fixation in search. Waldman et al. determined the average one glimpse probability for the whole search by averaging the one-glimpse probabilities of a rectangular grid of fixations over the whole search area. Their assumption of a grid-like fixation pattern with an equal probability of fixation is not supported by the literature on eye-movement research. The work of Noton and Stark [27] shows that people have repetitive and distinctive ways of inspecting and recognising a particular familiar object; the so-called scanpaths. The mathematical model of this thesis generates the fixation points in

sequence using Stark and Ellis' model [80] to predict eye movement fixation in a known scene. This model is true for targets and scenes with which the operator has a-priori knowledge. In this dissertation all the operators were first trained on the scene. This assumption is acceptable for the tests in this thesis, as all the operators will have a training session to familiarise themselves with the background and the targets. If a number of observers scan the same area, and their fixations are measured, it is found that the search area has certain regions of interest on which fixations occur repeatedly. If these regions are defined as states into which the fixations must be located, and we postulate that the transitions from one state to the other have a certain probability, we can effectively describe the generating process for these sequence of fixations as a Markov process. That is:

$$\bar{\mathbf{F}}_1 = \mathbf{M}_T \bar{\mathbf{F}}_0 \tag{A.8}$$

where $\mathbf{M_T}$ is the Markov probability transition matrix and $\mathbf{F}$ is the fixation state. The Markov model determines the probability of the next fixation state, which is realised by the drawing of a sample from a probability distribution. The inter-fixation angles follow from the state transitions, as each fixation state corresponds to a specific fixation area. The model is thus a statistical model and has to be repeated for a number of times as for a Monte Carlo model before meaningful results can be achieved.

The temporal model determines the glimpse time. Previous models often used a single value estimate for glimpse time, Bailey [21] used 0.33 s and Bloomfield [94] 0.25s. There is ample experimental evidence that glimpse time is a variable that depends on the size of the area to be searched and on the amount of clutter in the scene. Waldman models the fixation time as a modified power law:

$$t_g = (0.5782 + C_L)\theta_s^{-0.2132} \tag{A.9}$$

where C_L is the clutter factor and θ_s is the search sector in degrees. C_L depends on the similarity in size, general shape, and orientation between the background texture and the target, as well as on the background texture density. The possible values of C_L range from zero for a uniform background to unity for a background completely packed with elements of the same size, shape and orientation as the target. To the human eye C_L =0.05 looks like light clutter, C_L =0.15 looks like moderate clutter and C_L =0.25 like heavy clutter.

A.8 DETECTION SUB-MODEL

The model description starts with Seyb's [96] single glimpse probability of detecting a target in one specified fixation:

$$P_S = 1 - 0.5^{(F_H C_R)^3} \tag{A.10}$$

with

$$C_R \equiv \frac{C}{C_T} \tag{A.11}$$

where C is the displayed target contrast,

C_T the contrast threshold of the eye, and

F_H is the human factor (≥0 and ≤1), that will be discussed later.

The work of Blackwell [91] was used as described below to determine the contrast threshold. The foveal thresholds are generated from curve fits to Blackwell's 1946 data, keeping the full dependence on display luminance and displayed target angular size. The data naturally breaks into a high light level region and a low light region that is designated photopic and scotopic respectively, with the division at about 1.5×10^{-4} nt. In the photopic region the main receptors used by the eye are the cones while in the scotopic region rods are dominant. Waldman's curve fits to Blackwell's numbers are given for Photopic foveal vision as:

$$\log C_{FT} = (0.075 - 1.48 \times 10^{-4}\alpha) \cdot (\log 0.2919L - 2.96 + 1.025 \log \alpha)^2 - 0.601 \log \alpha - 1.04 \tag{A.12}$$

and Scotopic as:

$$\log C_{FT} = 0.1047(\log 0.2919L - 2.96 + 1.025 \log \alpha)^2 - 1.64 \log \alpha + 1.823 \tag{A.13}$$

where L is the luminance in nit (nt or cd/m^2), C_{FT} is the foveal contrast threshold and α is the target angular subtense in minutes of arc.

For the parafoveal case it is necessary to introduce another variable, θ, that is the angular offset of the target from the visual axis. Wootton's model (equation 13 in [102]) for the effect of the target in the peripheral vision on the contrast threshold is used:

$$C_{PFT} = C_{FT}\left[1 + 0.5\left(1 + \frac{\log 0.2919L}{3}\right)\theta^{(1.5+0.1 \log 0.2919L)}\right] \tag{A.14}$$

Where C_{PFT} is the parafoveal (peripheral) contrast threshold. The C_T term of equation (A.11) follows from C_{FT} (A.13 and A.14) or C_{PFT} (A.14), depending on whether the target is in the foveal or peripheral area of the eye.

The output of the human sub-model is used to drive the fixation sub-model and the detection sub-model. The single glimpse probabilities are accumulated in time over the allocated search time, and the detection probability is determined. The number of glimpses is dependent on the search time:

$$N_{\text{glimpses}} = \frac{T_{search}}{t_g} \tag{A15}$$

and

$$P_{\text{det}} = 1 - \left(1 - P_g\right)^{N_{\text{glimpses}}} \tag{A16}$$

A.9 RECOGNITION AND ACQUISITION MODEL

Finally, the recognition probability is determined using Johnson's criteria as refined in the RCA handbook [20]

$$P_{rec} = 1 - \exp\left(-0.856 \left(\frac{N_r - 0.4}{4} \right)^2 \right) \tag{A17}$$

where N_r is the number of resolution cells in the minimum target dimension of the system.

The target acquisition probability is determined from equation (1.1). For this dissertation it is assumed that there is no line of sight obstruction of the target, so P_{los} in equation (1.1) is always unity:

$$P_{acq} = P_{\text{det}} \times P_{rec} \tag{A.18}$$

APPENDIX B

TARGET ACQUISITION MODEL LISTINGS

B.1. NOMINAL VISDET LISTING

The VISDET model has been implemented in Matlab. The programme listing is given here.

```
% ************************************************************************
%                                                                        *
%  Program          : VDnom.M                                            *
%                                                                        *
%  Programmer       : G T Viljoen    3/11/95                             *
%                                                                        *
%  Purpose          : Program to implement Waldman's Visdet model to     *
%                   : determine target acquisition probability           *
%                                                                        *
% ************************************************************************

%  Scene parameters

STime=30.0;                  % Search time [s]
CL=0.04;                     % Clutter factor
LOB=300.0;                   % Background luminance [fL]
S=1.5;                       % Sky-ground luminance ratio
TgtSize=3.0;                 % Average target dimension [m]
TgtRefl=0.08;                % Target reflectance
BackRefl=0.1;                % Background reflectance
Saz=20.;Sel=5.;              % Search area, azimuth and elevation [deg]

%  Atmospheric parameters

Rvis=12000.0;                % Visibility range [m]
BIC=1.0;                     % Battlefield induced contaminants factor
Ptype=1;                     % Precipitation type 1-rain, 2-sleet, 3-snow
r=0.0;                       % Precipitation rate [mm/h]


% ************************************************************************
%                          VISDET MODEL                                  *
% ************************************************************************

%  Scene submodel

C0=abs(TgtRefl-BackRefl);

%  Atmospheric submodel

ii=0;
for R=1:200:4001,                       % Range to target [m]
  ii=ii+1;Rtgt(ii)=R;
  tau1 = exp(-3.91*R/Rvis);             % Aerosol scattering transmittance
  if Ptype==1, a=0.248*r^0.67; end      % Rain
  if Ptype==2, a=0.32*r^0.91; end       % Sleet (needle snow cristals)
  if Ptype==3, a=1.3*r^0.5; end         % Snow (Large flakes)
  tau2 = exp(-a*R);                     % Precipitation scattering transmittance
  TauAtm = tau1*tau2*BIC;               % Total Atmospheric transmittance
  CR = C0/( 1 - S*(1-1/TauAtm) );       % Propagated Contrast at range R
  LRB = TauAtm*LOB*C0/CR;               % Propagated background luminance at R

%  Sensor and display model

  alpha = 60.*57.296*TgtSize/R;         % Target angular subtense [minutes]
  C=CR;                                 % Human eye as sensor
  L=LRB;

%  Human submodel

  tg=(0.5782+CL)*Saz^(-0.2132);         % Average glimpse time [s]
```

```
  s = 0.152*tg^(-3.127);                % Average saccade amplitude [deg]
  m=2*round((Saz/(2*s))-0.5)+1;         % Define search block azimuth
  n=2*round((Sel/(2*s))-0.5)+1;         % and elevation
  loga=log10(alpha);
  tmp=(log10(L)-2.96+1.025*loga);
  LogCft=(0.075-0.000148*alpha)*tmp*tmp - 0.601*loga - 1.04;
  Cft = 10^LogCft;                      % Contrast threshold
  for j=1:m,
    for k=1:n,
      ps =  s*(j - 1 - (m-1)/2.); qs = s*(k - 1 - (n-1)/2.);
      the = sqrt(ps*ps + qs*qs);
      Cpft = Cft*(1 + 0.5*(1+log10(L)/3.)*the^(1.5+0.1*log(L)));
      Cratio = C/Cpft;
      FH = exp(-5.46*(CL^2.37));       % Human factor = f(Clutter);
      Ps(j,k)=1-0.5^((FH*Cratio)^3.0);
    end
  end
  Pg = sum(sum(Ps))/(m*n);
  tr=0.106*exp(3.57*(CL^1.8));
  nglimpse = (STime-tr)/tg;
  Pd=1-(1-Pg)^nglimpse;

%  Target recognition probability

  Pr=1.0;                   % As nominal results are for detection only

%  Target acquisition probability

  Pacq(ii)=Pd*Pr;
end

Pref=[1.0 1.0 1.0 0.99 0.975 0.96 0.895 0.72 0.51 0.33 0.205 0.13 0.08 0.05 0.04 0.025 0.02
      0.01 0.01 0.005 0.005];

plot(Rtgt,[Pacq' Pref]);grid
xlabel('Range (m)'),ylabel('Detection Probability')
title('VISDET Acceptance')
text(Rtgt(13),Pacq(13),'Visdet')
text(Rtgt(7),Pref(7),'Reference')
```

B.2. DIRECT VISION TEST MODEL TO DETERMINE CLUTTER VALUES IN SCENE

```
% ************************************************************************
%                                                                        *
%  Program          : VDdvt.M                                            *
%                                                                        *
%  Programmer       : G T Viljoen   11/9/99                              *
%                                                                        *
%  Purpose          : Program to implement Waldman's Visdet model to     *
%                   : determine Clutter values for direct vision         *
%                   : target acquisition probability                     *
%                                                                        *
% ************************************************************************

%  Scene parameters

LOB=50;                         % Background luminance [fL]
S=1.5;                          % Sky-ground luminance ratio
TgtSizeAz=7.0;TgtSizeEl=2.1; % Target dimensions [m]
TgtSize=(TgtSizeAz+TgtSizeEl)/2.0;  % Average target dimension [m]
Saz=120.;Sel=12.;               % Search area, azimuth and elevation [deg]

%  Atmospheric parameters

Rvis=20000.0;                   % Visibility range [m]
BIC=1.0;                        % Battlefield induced contaminants factor
Ptype=1;                        % Precipitation type 1-rain, 2-sleet, 3-snow
r=0.0;                          % Precipitation rate [mm/h]

%  Sensor parameters

EyeRes=0.0003;                  % Eye limiting resolution [rad]
```

```
%  Human model parameters


% ************************************************************************
%                          VISDET MODEL                                   *
% ************************************************************************

%  Scene parameters variations

for C0=0.2:0.1:0.3,          % Target background contrast
  for R=250:250:1000,        % Target range [m]
    if R==250,               % Search time [s]
      STime=10.0;
    else
      STime=30.0;
    end

%  Atmospheric submodel

    tau1 = exp(-3.91*R/Rvis);              % Aerosol scattering transmittance
    if Ptype==1, a=0.248*r^0.67; end       % Rain
    if Ptype==2, a=0.32*r^0.91; end        % Sleet (needle snow cristals)
    if Ptype==3, a=1.3*r^0.5; end          % Snow (Large flakes)
    tau2 = exp(-a*R);                      % Precipitation scattering transmittance
    TauAtm = tau1*tau2*BIC;                % Total Atmospheric transmittance
    CR = C0/( 1 - S*(1-1/TauAtm) );        % Propagated Contrast at range R
    LRB = TauAtm*LOB*C0/CR;                % Propagated background luminance at R

%  Sensor and display model


    alpha = 60.*57.296*TgtSize/R;        % Target angular subtense [minutes]
    C=CR;                                % Displayed contrast, human eye as sensor
    L=LRB;                               % Human eye as sensor

%  Human submodel

    for ii=1:25,
      CL=0.04*ii; clutter(ii)=CL;
      tg=(0.5782+CL)*Saz^(-0.2132);       % Average glimpse time [s]
      s = 0.152*tg^(-3.127);              % Average saccade amplitude [deg]
      m=2*round((Saz/(2*s))-0.5)+1;       % Define search block azimuth
      n=2*round((Sel/(2*s))-0.5)+1;       % and elevation
      loga=log10(alpha);
      tmp=(log10(L)-2.96+1.025*loga);
      LogCft=(0.075-0.000148*alpha)*tmp*tmp - 0.601*loga - 1.04;
      Cft = 10^LogCft;                    % Contrast threshold
      for j=1:m,
        for k=1:n,
          ps =  s*(j - 1 - (m-1)/2.); qs = s*(k - 1 - (n-1)/2.);
          the = sqrt(ps*ps + qs*qs);
          Cpft = Cft*(1 + 0.5*(1+log10(L)/3.)*the^(1.5+0.1*log(L)));
          Cratio = C/Cpft;                % Contrast ratio
          FH = exp(-5.46*(CL^2.37));      % Human factor = f(Clutter);
          Ps(j,k)=1-0.5^((FH*Cratio)^3.0); % Single glimpse Pdetection
        end
      end
      Pg = sum(sum(Ps))/(m*n);            % Mean detection probability
      tr=0.106*exp(3.57*(CL^1.8));        % Eye reaction time
      nglimpse = (STime-tr)/tg;           % Number of glimpses
      Pd=1-(1-Pg)^nglimpse;               % Detection probability

%  Target recognition probability

      Nr=TgtSizeEl/(R*EyeRes);            % Number of resolution cells over
                                          % minimum target dimension
      Pr=1-exp(-0.856*((Nr-0.4)/4)^2.0);  % Bailey's Recognition probability

%  Total target acquisition probability

      Pacq(ii)=Pd*Pr;
    end

% Plot results

%    plot(clutter,Pacq,'b');grid
%    xlabel('Clutter Factor'),ylabel('Detection Probability')
```

```
%    title('DVT Clutter Calibration')

    if C0==0.2,
      if R==250,p250c20=Pacq,end
      if R==500,p500c20=Pacq,end
      if R==750,p750c20=Pacq,end
      if R==1000,p1000c20=Pacq,end
    else
      if R==250,p250c30=Pacq,end
      if R==500,p500c30=Pacq,end
      if R==750,p750c30=Pacq,end
      if R==1000,p1000c30=Pacq,end
    end

  end
end

save dvtclut clutter p250c20 p500c20 p750c20 p1000c20 p250c30 p500c30 p750c30 p1000c30
```

The direct vision clutter data was plotted with clutplot.m to give figures 5.15 and 5.16:

```
% ************************************************************************
%                                                                        *
%  Program          : Clutplt.M                                          *
%                                                                        *
%  Programmer       : G T Viljoen   11/9/99                              *
%                                                                        *
%  Purpose          : Program to plot figures 5.14 and 5.15 in Thesis    *
%                   : and provides clutter data for table 5.7 and        *
%                   : vdesfid.m models                                   *
%                                                                        *
% ************************************************************************

% Data saved in dvtclut
load dvtclut

% Plot DVT Acquisition probability as a function of clutter value

figure(1)
plot(clutter,[p250c20' p500c20' p750c20' p1000c20']);grid;
xlabel('Clutter Factor'),ylabel('Acquisition Probability')
title('DVT Clutter Calibration, 20% contrast targets')
text(clutter(9),p250c20(9),'250m');
text(clutter(12),p500c20(12),'500m');
text(clutter(12),p750c20(12),'750m');
text(clutter(12),p1000c20(12),'1000m');

figure(2)
plot(clutter,[p250c30' p500c30' p750c30' p1000c30']);grid;
xlabel('Clutter Factor'),ylabel('Acquisition Probability')
title('DVT Clutter Calibration, 30% contrast targets')
text(clutter(10),p250c30(10),'250m');
text(clutter(12),p500c30(12),'500m');
text(clutter(12),p750c30(12),'750m');
text(clutter(12),p1000c30(12),'1000m');

% Make clutter functions monotone for interpolation purposes

for i=1:max(size(clutter))-1,
  if(abs(p250c20(i+1)-p250c20(i))<=0.0001)|(p250c20(i+1)>p250c20(i)),
    p250c20(i+1)=p250c20(i)-0.0001;
  end
  if(abs(p500c20(i+1)-p500c20(i))<=0.0001)|(p500c20(i+1)>p500c20(i)),
    p500c20(i+1)=p500c20(i)-0.0001;
  end
  if(abs(p750c20(i+1)-p750c20(i))<=0.0001)|(p750c20(i+1)>p750c20(i)),
    p750c20(i+1)=p750c20(i)-0.0001;
  end
  if(abs(p1000c20(i+1)-p1000c20(i))<=0.0001)|(p1000c20(i+1)>p1000c20(i)),
    p1000c20(i+1)=p1000c20(i)-0.0001;
  end
  if(abs(p250c30(i+1)-p250c30(i))<=0.0001)|(p250c30(i+1)>p250c30(i)),
    p250c30(i+1)=p250c30(i)-0.0001;
  end
```

```
  if(abs(p500c30(i+1)-p500c30(i))<=0.0001)|(p500c30(i+1)>p500c30(i)),
    p500c30(i+1)=p500c30(i)-0.0001;
  end
  if(abs(p750c30(i+1)-p750c30(i))<=0.0001)|(p750c30(i+1)>p750c30(i)),
    p750c30(i+1)=p750c30(i)-0.0001;
  end
  if(abs(p1000c30(i+1)-p1000c30(i))<=0.0001)|(p1000c30(i+1)>p1000c30(i)),
    p1000c30(i+1)=p1000c30(i)-0.0001;
  end
end

% Do interpolation of clutter curves to get clutter factor for each target
%      1    2    3    4    5    6    7    8    9    10   11   12   13   14   15   16
p20=[0.77 0.31 0.46 0.38 0.38 0.54 0.77 0.77 0.15 0.85 0.38 0.92 0.38 0.97 0.77 0.38];
p30=[0.83 0.42 0.25 0.25 0.33 0.42 0.83 0.97 0.25 0.97 0.33 0.83 0.42 0.97 0.58 0.42];
r=[500 750 250 750 500 1000 250 500 1000 250 750 500 1000 750 250 1000];
for i=1:16,                          % For each target
  if r(i)==250,
    a=[p250c20' clutter'];cl(i,1)=table1(a,p20(i));
    a=[p250c30' clutter'];cl(i,2)=table1(a,p30(i));
  end
  if r(i)==500,
    a=[p500c20' clutter'];cl(i,1)=table1(a,p20(i));
    a=[p500c30' clutter'];cl(i,2)=table1(a,p30(i));
  end
  if r(i)==750,
    a=[p750c20' clutter'];cl(i,1)=table1(a,p20(i));
    a=[p750c30' clutter'];cl(i,2)=table1(a,p30(i));
  end
  if r(i)==1000,
    a=[p1000c20' clutter'];cl(i,1)=table1(a,p20(i));
    a=[p1000c30' clutter'];cl(i,2)=table1(a,p30(i));
  end
end

% Print clutter values to be copied to vdesfid.m

cl
```

B.3. ESFID MODEL WITH GRID FIXATION PATTERN

```
% ************************************************************************
%                                                                        *
%  Program          : VDesfid1.M                                         *
%                                                                        *
%  Programmer       : G T Viljoen   12/9/99                              *
%                                                                        *
%  Purpose          : Program to implement Waldman's Visdet model to     *
%                   : determine ESFID target acquisition probability     *
%                   : This model uses Waldman's grid for eye fixations   *
%                                                                        *
% ************************************************************************

%  Scene parameters for each target

C0=0.3;                        % Target background contrast
% Target range [m]
range=[500 750 250 750 500 1000 250 500 1000 250 750 500 1000 750 250 1000];
% Search time [s]
searchtime=[30 30 10 30 30 30 10 30 30 10 30 30 30 30 10 30];
% Clutter model from direct vision calibration model vddvt.m and clutplt.m
% clutter=[
% 0.42 0.56 0.51 0.54 0.58 0.45 0.3  0.42 0.66 0.21 0.54 0.2  0.54 0.2 0.3  0.54
% 0.51 0.59 0.65 0.65 0.66 0.54 0.41 0.2  0.62 0.2  0.63 0.51 0.54 0.2 0.52 0.54]';
clutter=[
    0.4802    0.5066
    0.5846    0.5975
    0.5105    0.6491
    0.5368    0.6398
    0.5990    0.6565
    0.4553    0.5231
    0.2648    0.4501
    0.4802    0.2680
    0.6290    0.6166
    0.2524    0.2420
```

```
    0.5368    0.6204
    0.2648    0.5066
    0.5076    0.5231
    0.2434    0.2457
    0.2648    0.5251
    0.5076    0.5231];

LOB=50;                         % Background luminance [fL]
S=1.5;                          % Sky-ground luminance ratio
TgtSizeAz=7.0;TgtSizeEl=2.1; % Target dimensions [m]
TgtSize=(TgtSizeAz+TgtSizeEl)/2.0;  % Average target dimension [m]
Saz=120.;Sel=12.;               % Search area, azimuth and elevation [deg]

%  Atmospheric parameters

Rvis=20000.0;                   % Visibility range [m]
BIC=1.0;                        % Battlefield induced contaminants factor
Ptype=1;                        % Precipitation type 1-rain, 2-sleet, 3-snow
r=0.0;                          % Precipitation rate [mm/h]

%  Sensor parameters

Nmonitor=1080;                  % Number of active lines on ESFID monitors
NFOV=(1000.0/57.296)*9.28;     % Vertical field of view of NF monitor [mrad]
WFOV=(1000.0/57.296)*46.4;     % Vertical field of view of WF monitor [mrad]
AR=1.777;                       % Aspect ratio of monitor
sigmon=1/(2.31*Nmonitor);
LmonMax=75.0;                   % Maximum monitor luminance [fL](75)
MonRes=NFOV/(1000.0*Nmonitor); % Monitor limiting resolution [rad] for recognition (0.488mrad
nom)

%  Human model parameters

Ncr=4.0                         % Number of cycles required on the target for recognition

% ************************************************************************
%                          VISDET MODEL                                   *
% ************************************************************************

for ii=1:16,                    % For each target
%  Scene submodel

  R = range(ii);                        % Target range [m]
  STime=searchtime(ii);                 % Search time [s]
  if C0==0.2, jj=1; else, jj=2; end
  CL = clutter(ii,jj);                  % Scene contrast

%  Atmospheric submodel

  tau1 = exp(-3.91*R/Rvis);             % Aerosol scattering transmittance
  if Ptype==1, a=0.248*r^0.67; end      % Rain
  if Ptype==2, a=0.32*r^0.91; end       % Sleet (needle snow cristals)
  if Ptype==3, a=1.3*r^0.5; end         % Snow (Large flakes)
  tau2 = exp(-a*R);                     % Precipitation scattering transmittance
  TauAtm = tau1*tau2*BIC;               % Total Atmospheric transmittance
  CR = C0/( 1 - S*(1-1/TauAtm) );       % Propagated Contrast at range R
  LRB = TauAtm*LOB*C0/CR;               % Propagated background luminance at R

%  Sensor and display model


  alpha = 60.*57.296*TgtSize/R;         % Target angular subtense [minutes]
  frnf=Ncr*NFOV*R/(1000.0*TgtSizeEl);   % Raster frequency on NF monitor [cycles]
  MTFnf=exp(-2*(pi*sigmon*frnf)^2.0);   % MTF of NF monitor for minimum target dimension
  frwf=Ncr*WFOV*R/(1000.0*TgtSizeEl);   % Raster frequency on WF monitor
  MTFwf=exp(-2*(pi*sigmon*frwf)^2.0);   % MTF of WF monitor for minimum target dimension
%  C=CR*MTFf;                           % Displayed contrast on NF CRT
  if LRB>LmonMax,                       % CRT's maximum luminance
    L=LmonMax;
  else,
    L=LRB;
  end

%  Human submodel

  tg=(0.5782+CL)*Saz^(-0.2132);         % Average glimpse time [s]
  s = 0.152*tg^(-3.127);                % Average saccade amplitude [deg]
```

```
  m=2*round((Saz/(2*s))-0.5)+1;          % Define search block azimuth
  n=2*round((Sel/(2*s))-0.5)+1;          % and elevation
  loga=log10(alpha);
  tmp=(log10(L)-2.96+1.025*loga);
  LogCft=(0.075-0.000148*alpha)*tmp*tmp - 0.601*loga - 1.04;
  Cft = 10^LogCft;                       % Contrast threshold
  for j=1:m,
    for k=1:n,
      ps =  s*(j - 1 - (m-1)/2.); qs = s*(k - 1 - (n-1)/2.); % Angle between fixation and
target [deg]
      the = sqrt(ps*ps + qs*qs);
      Cpft = Cft*(1 + 0.5*(1+log10(L)/3.)*the^(1.5+0.1*log(L)));
      if (ps<57.296*NFOV/2000.0)&(qs<57.296*NFOV*AR/2000.0),
        C=CR*MTFnf;                      % Displayed contrast on NF CRT
      else
        C=CR*MTFwf;                      % Displayed contrast on WF CRT
      end

      Cratio = C/Cpft;                   % Contrast ratio
      FH = exp(-5.46*(CL^2.37));         % Human factor = f(Clutter);
      Ps(j,k)=1-0.5^((FH*Cratio)^3.0); % Single glimpse Pdetection
    end
  end
  Pg = sum(sum(Ps))/(m*n);               % Mean detection probability
  tr=0.106*exp(3.57*(CL^1.8));           % Eye reaction time
  nglimpse = (STime-tr)/tg;              % Number of glimpses
  Pd=1-(1-Pg)^nglimpse;                  % Detection probability

%  Target recognition probability

  Nr=TgtSizeEl/(R*MonRes);               % Number of resolution cells over
                                         % minimum target dimension
  Pr=1-exp(-0.856*((Nr-0.4)/4)^2.0);     % Bailey's Recognition probability

%  Total target acquisition probability

  Pacq(ii)=Pd*Pr;
end

% Summarise results

Rplot=[250 500 750 1000];
Pplot(1)=(Pacq(3)+Pacq(7)+Pacq(10)+Pacq(15))/4.0;     % Mean 250m Pacq
Pplot(2)=(Pacq(1)+Pacq(5)+Pacq(8)+Pacq(12))/4.0;      % Mean 500m Pacq
Pplot(3)=(Pacq(2)+Pacq(4)+Pacq(11)+Pacq(14))/4.0;     % Mean 750m Pacq
Pplot(4)=(Pacq(6)+Pacq(9)+Pacq(13)+Pacq(16))/4.0;     % Mean 1000m Pacq

% Experimental results to compare with model

PRef=[0.5    0.55               % Experimental Pacq
      0.575  0.575
      0.25   0.425
      0.05   0.175];
nref=40;                         % Number of experimental subjects per data point
for i=1:4,                       % Standard deviation from binomial distribution
  for j=1:2,
    SigRef(i,j)=sqrt(nref*PRef(i,j)*(1-PRef(i,j)))/nref;
  end
end

% Plot data

if C0==0.2,
  plot(Rplot,[Pplot' PRef(:,1)]);grid;hold on
  errorbar(Rplot,PRef(:,1),SigRef(:,1));hold off
  xlabel('Range [m]'),ylabel('Acquisition Probability')
  title('ESFID Pacq for 20% Contrast targets')
  text(Rplot(2),Pplot(2),'Model');
  text(Rplot(1),PRef(1,1),'Experiment');
else
  plot(Rplot,[Pplot' PRef(:,2)]);grid;hold on
  errorbar(Rplot,PRef(:,2),SigRef(:,2));hold off
  xlabel('Range [m]'),ylabel('Acquisition Probability')
  title('ESFID Pacq for 30% Contrast targets')
  text(Rplot(2),Pplot(2),'Model');
  text(Rplot(1),PRef(1,2),'Experiment');
end
```

B.4. ESFID MODEL WITH MARKOV FIXATION PATTERN

```
% ************************************************************************
%                                                                        *
%  Program          : VDesfid2.M                                         *
%                                                                        *
%  Programmer       : G T Viljoen   12/9/99                              *
%                                                                        *
%  Purpose          : Program to implement Waldman's Visdet model to     *
%                   : determine ESFID target acquisition probability     *
%                   : This model uses Stark's Markov model to predict    *
%                   : eye fixations.                                     *
%                                                                        *
% ************************************************************************

%  Scene parameters for each target

C0=0.3;                         % Target background contrast
% Target range [m]
range=[500 750 250 750 500 1000 250 500 1000 250 750 500 1000 750 250 1000];
% Search time [s]
searchtime=[30 30 10 30 30 30 10 30 30 10 30 30 30 30 10 30];
% Clutter
clutter=[
    0.4802    0.5066
    0.5846    0.5975
    0.5105    0.6491
    0.5368    0.6398
    0.5990    0.6565
    0.4553    0.5231
    0.2648    0.4501
    0.4802    0.2680
    0.6290    0.6166
    0.2524    0.2420
    0.5368    0.6204
    0.2648    0.5066
    0.5076    0.5231
    0.2434    0.2457
    0.2648    0.5251
    0.5076    0.5231];
LOB=75;                         % Background luminance [fL]
S=1.5;                          % Sky-ground luminance ratio
TgtSizeAz=7.0;TgtSizeEl=2.1; % Target dimensions (7x2.1)[m]
TgtSize=(TgtSizeAz+TgtSizeEl)/2.0;  % Average target dimension [m]
Saz=120.;Sel=12.;               % Search area, azimuth and elevation [deg], 120/12
% Target centre coordinates in pixels on the scene
tp=[   5108.00       1239.00
        928.50       1171.50
       6951.50       1430.00
       5713.00       1227.50
       6608.50       1428.00
       2579.50       1193.50
        242.50       1432.00
       1855.00       1230.00
       6810.50       1231.50
       2890.00       1451.50
       7450.50       1243.50
       3334.00       1270.00
       5349.50       1190.50
       3279.50       1226.50
       5312.50       1410.50
        448.50       1147.50];
pscl=0.000244;                  % Pixel scale [rad/pixel]

%  Atmospheric parameters

Rvis=20000.0;                   % Visibility range [m]
BIC=1.0;                        % Battlefield induced contaminants factor
Ptype=1;                        % Precipitation type 1-rain, 2-sleet, 3-snow
r=0.0;                          % Precipitation rate [mm/h]

%  Sensor parameters

Nmonitor=400;                  % Number of active lines on ESFID monitors
NFOV=(1000.0/57.296)*6.67;     % Vertical field of view of NF monitor [mrad]
```

```
WFOV=(1000.0/57.296)*37.5;    % Vertical field of view of WF monitor [mrad]
AR=4/3;                       % Aspect ratio of monitor
sigmon=1/(2.31*Nmonitor);
LmonMax=75.0;                  % Maximum monitor luminance [fL](75)
MonRes=NFOV/(1000.0*Nmonitor); % Monitor limiting resolution [rad] for recognition (0.488mrad
nom)

%  Human model parameters

nsub=40.0                     % Number of subjects modelled
Ncr=4.0                       % Number of cycles required on the target for recognition
load markov                   % Markov M1 matrix to predict fixations
load peaks                    % The coordinates of the fixations
fj=193;                       % Start fixation position
rand('seed',200);             % Fix the seed to give repeatable results
np=max(size(pk2));            % Number of fixation positions

% ************************************************************************
%                          VISDET MODEL                                  *
% ************************************************************************

for ii=1:16,                  % For each target
%  Scene submodel

  R = range(ii);                          % Target range [m]
  STime=searchtime(ii);                   % Search time [s]
  if C0==0.2, jj=1; else, jj=2; end       % Scene contrast determines
  CL = clutter(ii,jj);                    % Scene clutter

%  Atmospheric submodel

  tau1 = exp(-3.91*R/Rvis);               % Aerosol scattering transmittance
  if Ptype==1, a=0.248*r^0.67; end        % Rain
  if Ptype==2, a=0.32*r^0.91; end         % Sleet (needle snow cristals)
  if Ptype==3, a=1.3*r^0.5; end           % Snow (Large flakes)

  tau2 = exp(-a*R);                       % Precipitation scattering transmittance
  TauAtm = tau1*tau2*BIC;                 % Total Atmospheric transmittance
  CR = C0/( 1 - S*(1-1/TauAtm) );         % Propagated Contrast at range R
  LRB = TauAtm*LOB*C0/CR;                 % Propagated background luminance at R

%  Sensor and display model

  alpha = 60.*57.296*TgtSize/R;           % Target angular subtense [minutes]
  frnf=Ncr*NFOV*R/(1000.0*TgtSizeEl);     % Raster frequency on NF monitor [cycles]
  MTFnf=exp(-2*(pi*sigmon*frnf)^2.0);     % MTF of NF monitor for minimum target dimension
  frwf=Ncr*WFOV*R/(1000.0*TgtSizeEl);     % Raster frequency on WF monitor
  MTFwf=exp(-2*(pi*sigmon*frwf)^2.0);     % MTF of WF monitor for minimum target dimension
  if LRB>LmonMax,                         % CRT's maximum luminance
    L=LmonMax;
  else,
    L=LRB;
  end

%  Human submodel

  tg=(0.5782+CL)*Saz^(-0.2132);           % Average glimpse time [s]
  tr=0.106*exp(3.57*(CL^1.8));            % Eye reaction time
  loga=log10(alpha);
  tmp=(log10(L)-2.96+1.025*loga);
  LogCft=(0.075-0.000148*alpha)*tmp*tmp - 0.601*loga - 1.04;
  Cft = 10^LogCft;                        % Contrast threshold for target (Blackwell)
  Psub=zeros(nsub,1);                     % Initialise probability per subject
  for i=1:nsub,                           % Model the number of subjects
    Time=0;k=0;Pc=0;                      % Initialise search time and cumulative Pd
    while Time<STime,                     % Target search model
      Time=Time+tg;k=k+1;                 % Time
      sr=rand;sm=0.0;                     % Draw random number to determine next fixation point
      for fi=1:np,                        % Markov fixation model
        sm = sm+M1(fj,fi);
        if sm>=sr,
          pk(k,:)=pk2(fi,:);
          fp(k)=fi;fj=fi;                 % Move to next fixation point
          break
        end
      end
      ps=tp(ii,1)-pk(k,1);qs=tp(ii,2)-pk(k,2); % Pixels between fixation and target
```

```
      the = 57.296*pscl*sqrt(ps*ps + qs*qs);   % Angle to target
      Cpft = Cft*(1 + 0.5*(1+log10(L)/3.)*the^(1.5+0.1*log(L)));
      thet=57.296*pscl*qs; psit=57.296*pscl*ps;
      if (thet<57.296*NFOV/2000.0)&(psit<57.296*NFOV*AR/2000.0),
        C=CR*MTFnf;                    % Displayed contrast on NF CRT
      else
        C=CR*MTFwf;                    % Displayed contrast on WF CRT
      end
      Cratio = C/Cpft;                 % Contrast ratio
      FH = exp(-5.46*(CL^2.37));       % Human factor = f(Clutter);
      Ps=1-0.5^((FH*Cratio)^3.0);      % Single glimpse Pdetection
      Pc=1-(1-Pc)*(1-Ps);              % Cumulative detection probability

    end
    Psub(i)=Pc;                        % Detection probability for each subject
    Ptel(ii,i)=Pc;                     % Store data in array for telemetry purposes
  end
  Pd = mean(Psub);                     % Average detection probability for subjects

%  Target recognition probability

  Nr=TgtSizeEl/(R*MonRes);                  % Number of resolution cells over minimum target
dimension
  Pr=1-exp(-0.856*((Nr-0.4)/4)^2.0);   % Bailey's Recognition probability

%  Total target acquisition probability

  Pacq(ii)=Pd*Pr;
end

% Summarise results

Rplot=[250 500 750 1000];
Pplot(1)=(Pacq(3)+Pacq(7)+Pacq(10)+Pacq(15))/4.0;     % Mean 250m Pacq
Pplot(2)=(Pacq(1)+Pacq(5)+Pacq(8)+Pacq(12))/4.0;      % Mean 500m Pacq
Pplot(3)=(Pacq(2)+Pacq(4)+Pacq(11)+Pacq(14))/4.0;     % Mean 750m Pacq
Pplot(4)=(Pacq(6)+Pacq(9)+Pacq(13)+Pacq(16))/4.0;     % Mean 1000m Pacq

% Experimental results to compare with model

PRef=[0.5    0.55              % ESFID measured Experimental Pacq
      0.575  0.575
      0.25   0.425
      0.05   0.175];
nref=40;                         % Number of experimental subjects per data point
for i=1:4,                       % Standard deviation from binomial distribution
  for j=1:2,
    SigRef(i,j)=sqrt(nref*PRef(i,j)*(1-PRef(i,j)))/nref;
  end
end

% Plot data

if C0==0.2,
  plot(Rplot,[Pplot' PRef(:,1)]);grid;hold on
  errorbar(Rplot,PRef(:,1),SigRef(:,1));hold off
  xlabel('Range [m]'),ylabel('Acquisition Probability')
  title('ESFID Pacq for 20% Contrast targets')
  text(Rplot(2),Pplot(2),'Model');
  text(Rplot(2),PRef(1,2),'Experiment');
else
  plot(Rplot,[Pplot' PRef(:,2)]);grid;hold on
  errorbar(Rplot,PRef(:,2),SigRef(:,2));hold off
  xlabel('Range [m]'),ylabel('Acquisition Probability')
  title('ESFID Pacq for 30% Contrast targets')
  text(Rplot(2),Pplot(2),'Model');
  text(Rplot(2),PRef(2,2),'Experiment');
end
```

APPENDIX C

DETAILED ESFID HARDWARE DESCRIPTION

C.1 HARDWARE DESCRIPTION

The hardware block diagram of the ESFID display is shown in figure C1. The ESFID display consists of the wide field display, the narrow-field display, the narrow-field projector optics, the eye tracker camera and LED, the eye tracker DSP, and the helmet tracker and the projector servo control. The interface of the ESFID display to the rest of the teleoperator system consists of two video cables carrying the wide and narrow-field video data, two RS422 interface cables with the eye tracker azimuth and elevation angle information, and one RS232 serial interface with the helmet tracker information.

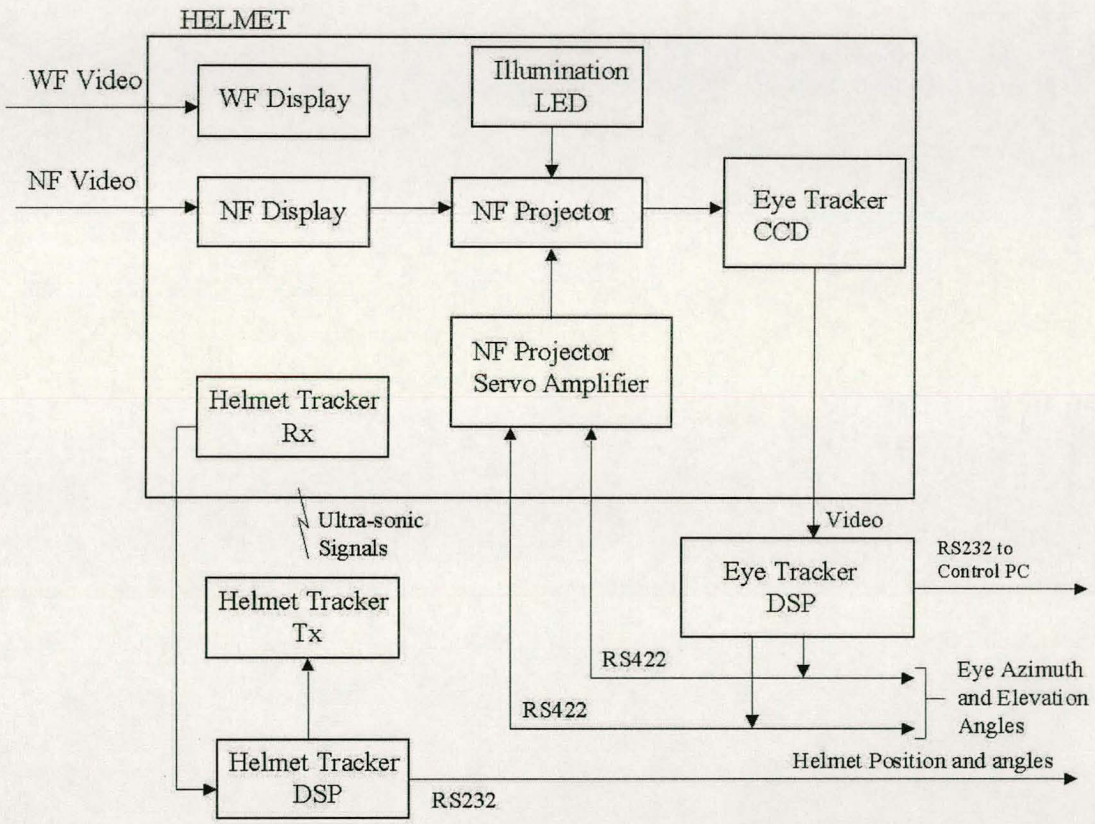


Figure C1 ESFID Hardware Block Diagram

The ESFID helmet display that was built is shown in figure C2, from a bottom view. It shows the wide field display, the narrow-field display, the NF projector optics and the eye tracker camera housing.

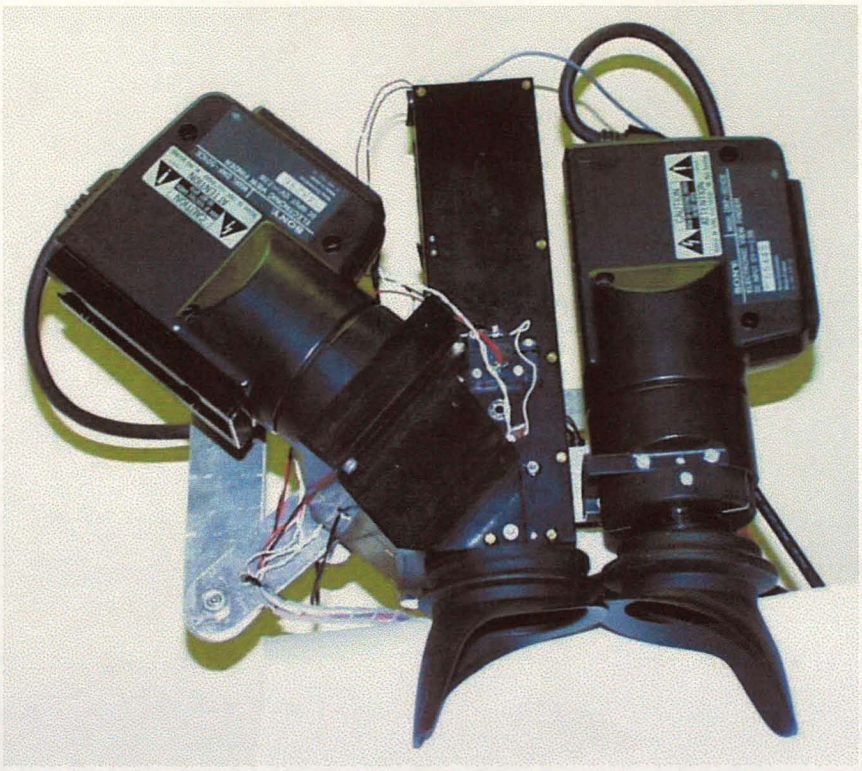


Figure C2 ESFID Bottom View

Figure C3 gives the front view of the ESFID. It is clear that the WF and NF channels are split, and the mechanics to adjust the displays for interpupillary distance, azimuth (NF) and elevation (WF) angles are shown.

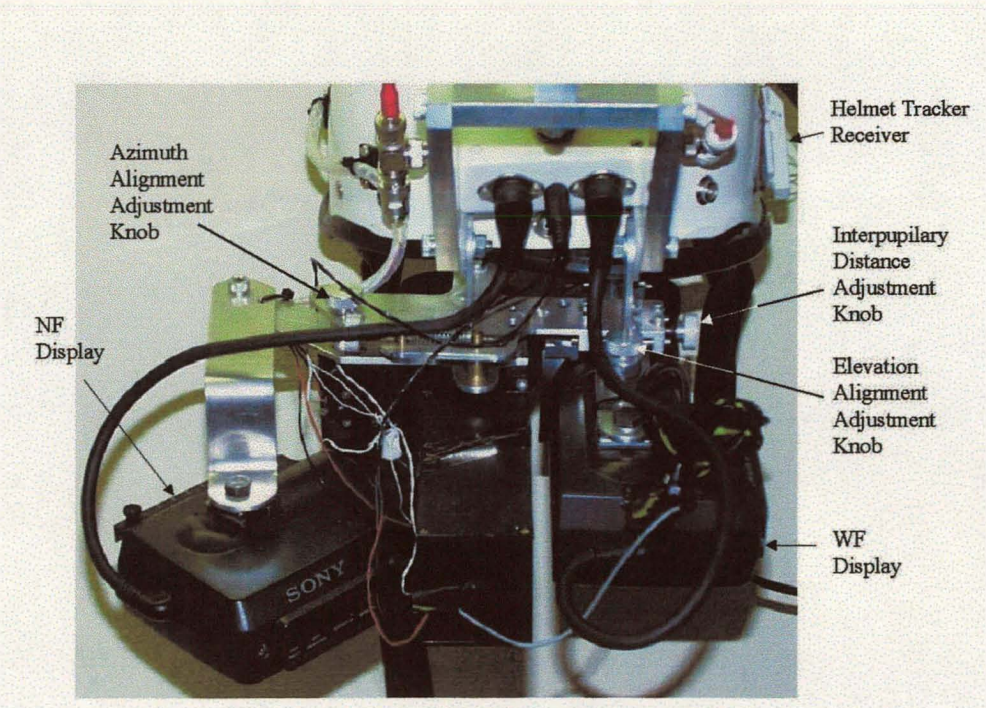


Figure C3 ESFID Front View

The ESFID side view shows the narrow field projector optics module and the eye piece. The mirror projection system is housed in the NF projector module. The view with the NF eye-piece removed is shown in figure C5. The dichroic azimuth mirror, together with the azimuth mirror gear and the reference LED is visible.

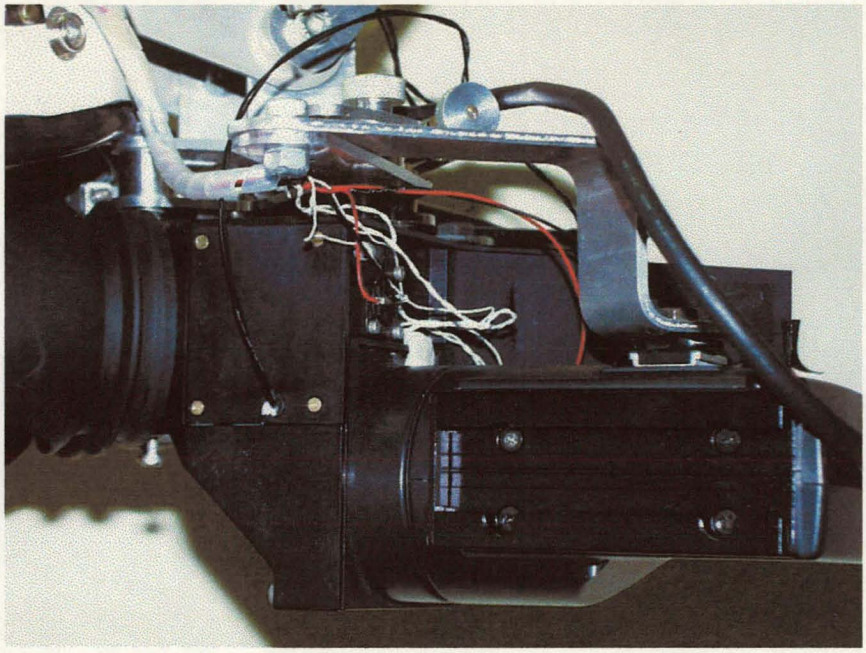


Figure C4 ESFID Side View, Showing NF Projector Unit

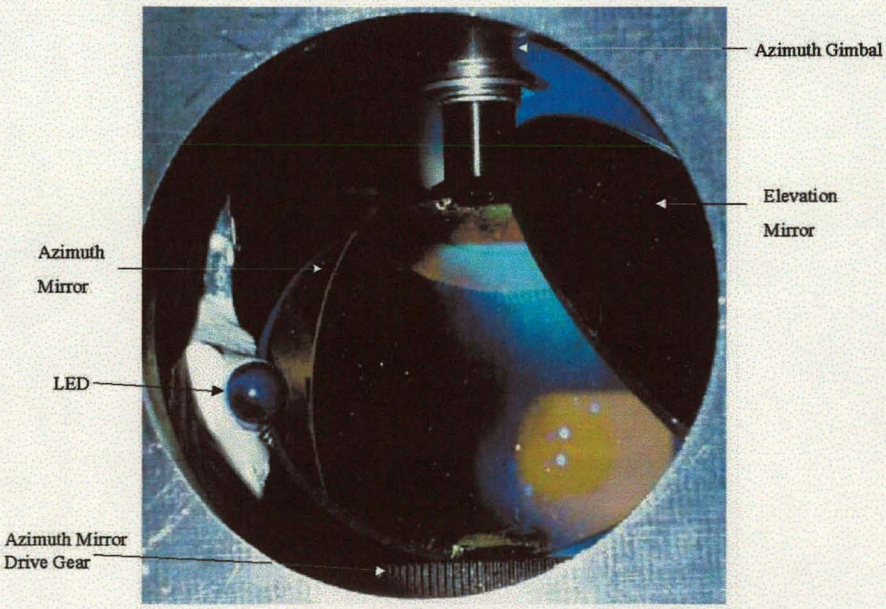


Figure C5 NF Azimuth Mirror and reference LED

Figure C6 gives another view into the NF projector through the eye-piece opening. In this view the elevation mirror is clearly visible, together with its drive gear in the background. The azimuth mirror and its drive gear are also indicated to aid orientation.

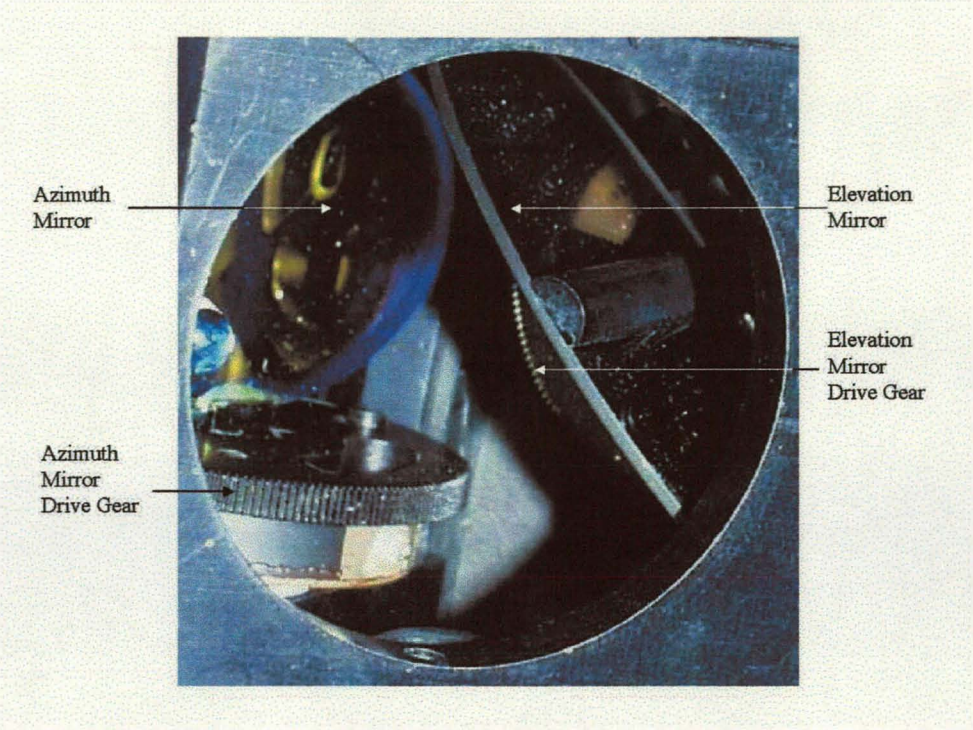


Figure C6 NF Elevation Mirror

The mirrors are driven with small electric motors through an analogue linear servo amplifier using potentiometer feedback. The 1W servo driver boards for each of the mirrors are mounted on the back of the helmet, to help counterbalance the front display. The servo driver boards are shown in figure C7.

The eye-tracking DSP, algorithms and software was done by B. Nortje [98]. The eye measurement camera was a single chip Peach camera that was mounted on the NF projector as shown in figure C8. The camera worked in the 650nm to 1000nm spectrum using illumination provided by the LED. The camera required only 12 V power, and the output was PAL video that was sent to the eye-tracker DSP box. The focus of the camera could be set with the adjustment screw in the back of the mounting.

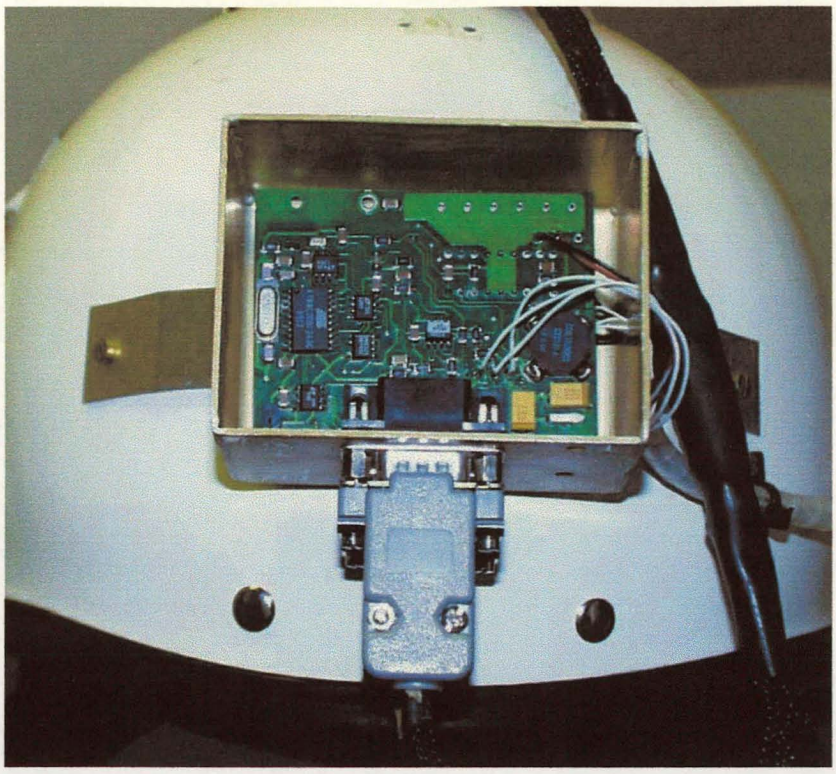


Figure C7 ESFID Back view, Showing Projector Servo Amplifiers

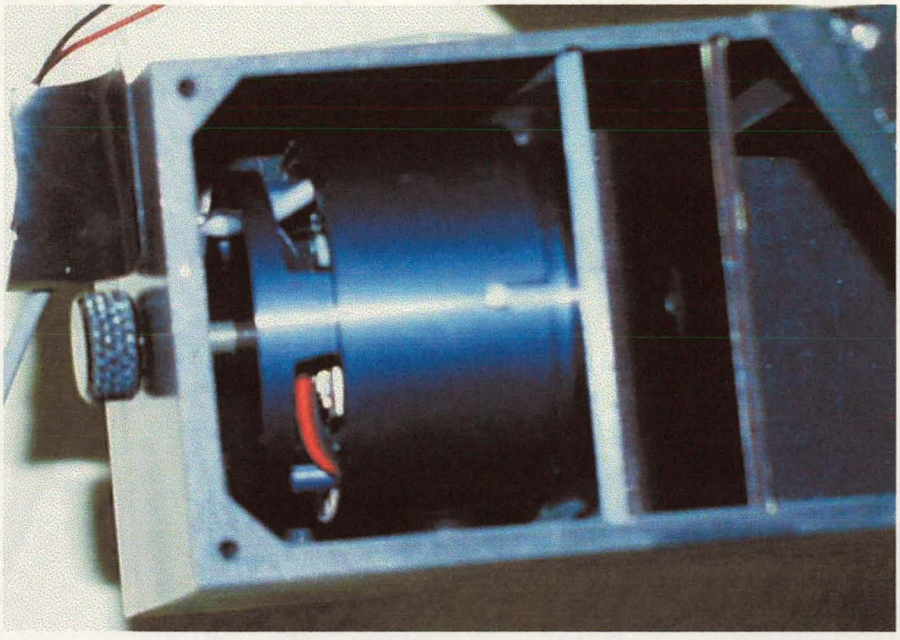


Figure C8 Eye tracker CCD Camera

The eye tracker DSP card is shown in figure C9. The board is a half Euro-card format and it is populated on both sides. The board includes a video frame grabber with a 10MHz 8-bit A/D, two RS422 outputs for the azimuth and elevation eye tracking error and a RS232 serial port. The digitised video is stored in three switchable banks

of video RAM. The one bank is used for the current frame that is being sampled, the second bank is used to do the eye-tracking processing on, and the third bank is used to generate a video output that includes the tracking cross-hairs. After each video frame, the banks are switched. The processor is a TMS 320C31 60 MHz DSP chip. The interface to the control PC is done through a C8051 processor and RS232 communications.

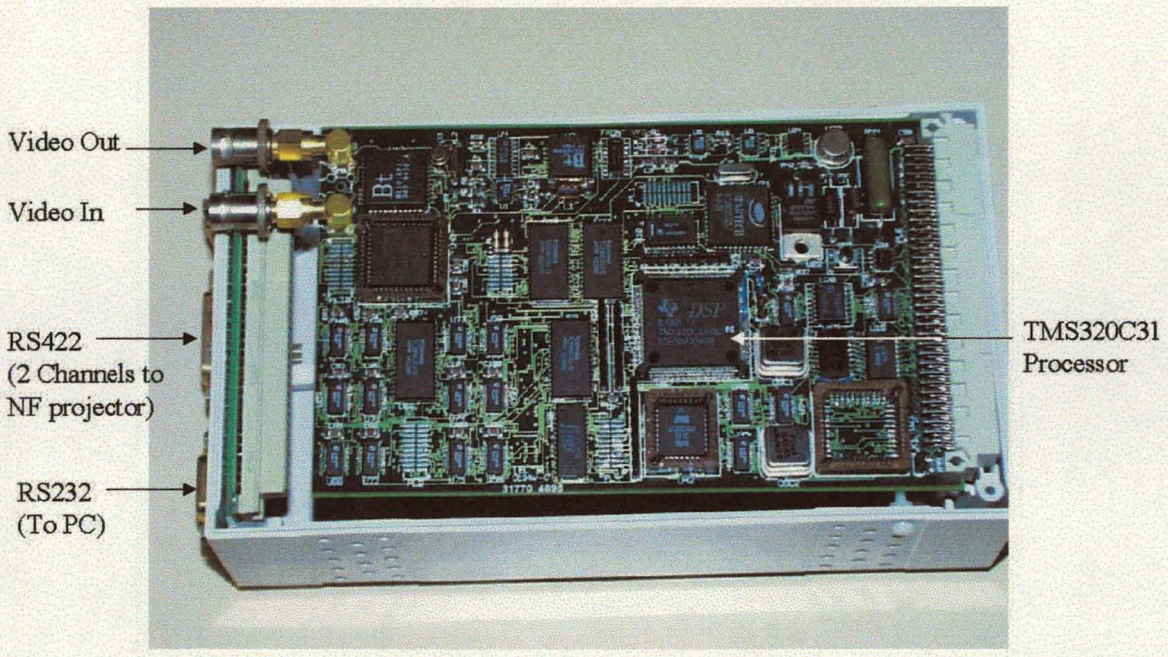


Figure C9 Eye Tracker DSP

The eye tracker DSP algorithms were developed through the 8051 interface. The eye tracker calibration and neural net training was also done through this interface. During operations in the ESFID simulator, the 8051 was not used.

C.2 ESFID USER DESCRIPTION

The test subjects each have a unique eye physiology, such as pupil size, eye distance from the tracker (the depth of the eye in the face), the position and size of the eye lids and eyelashes, and corneal radius. The eye-tracking subsystem of the ESFID system thus has to be calibrated for each test subject.

The flow diagram of the how the ESFID is operated is shown in figure C10. The subject is strapped into a jig, shown in figure 4.8, that keeps his eye position constant in the middle of the display throughout the calibration process. Although the eye-tracker concept is not very sensitive to eye-movement, the fact that the illumination LED's light was uncollimated, a small sensitivity to eye position remained. It was also important to keep the same illumination levels during the calibration process, as the eye-tracker camera was equipped with an automatic electronic IRIS that could influence the eye-tracking halfway through the calibration.

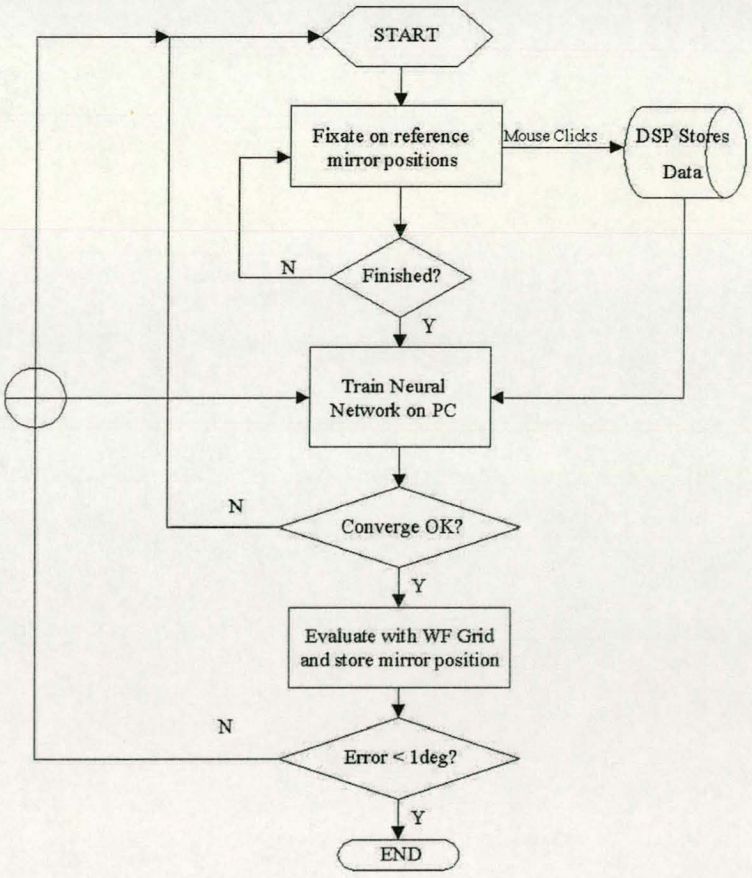


Figure C10 ESFID Operational Flow Diagram

The operator is presented with a flashing dot in the middle of the NF display, while the wide-field displays a uniform grey scene. The mirror is moved to a known position, by commanding the mirror servo with a 12bit word. The levels of the mirror servo command were the units used for the calibration. A pattern of 50 mirror positions is shown sequentially to the operator, and he is asked to fixate on the flashing dot for a period of 2-3s before he presses a mouse button on the control PC. The DSP is commanded by the control PC to record the tracking crosshairs for 10 samples, or 200ms. One calibration session gathers 500 data points. The difference between the tracking crosshairs, as described in figure 4.7, is then transmitted from the DSP to the control PC.

In the control PC, the neural network training algorithm is run on the calibration data. The training have to be done in the control PC due to memory requirements, the neural network training algorithm required more memory than was available on the DSP. Once the training is completed, the neural network weights for the subject is stored and downloaded again to the DSP hardware, where the neural network algorithm will use the weights for the real time tracking.

The final step is to evaluate the tracking performance of the eye-tracker. The operator is presented with a grid pattern with 2 degrees between the lines in his wide field of view display. The narrow-field display is a X crosshair. The operator is asked to align the crosshair on the grid by moving the wide-field grid one pixel at a time in various directions as indicated by the operator. The operator can then move the mirror with his eye and check if it follows his eye movements over a ±10° field of regard. The criteria is eye-tracking to within one degree accuracy, or about within ¼ of the vertical screen in the wide-field. Finally, an eye tracking accuracy measurement is done. The narrow-field display is switched off, and on the wide-field a large block is displayed with a flashing cross on one of the corners. The size of the block and the coordinates of the crosses are known. The operator is asked to fixate on the flashing cross for a period of 5s, while measurements are recorded of the mirror movement. This process is repeated over 9 points, as shown in figure C11. The mirror position error is then plotted in Matlab, to give a standard deviation tracking error and a mean tracking error. If the tracking is not satisfactory, either the neural network is retrained, or the whole calibration is repeated. It was found that most subjects had to be calibrated twice before they achieved consistent results.

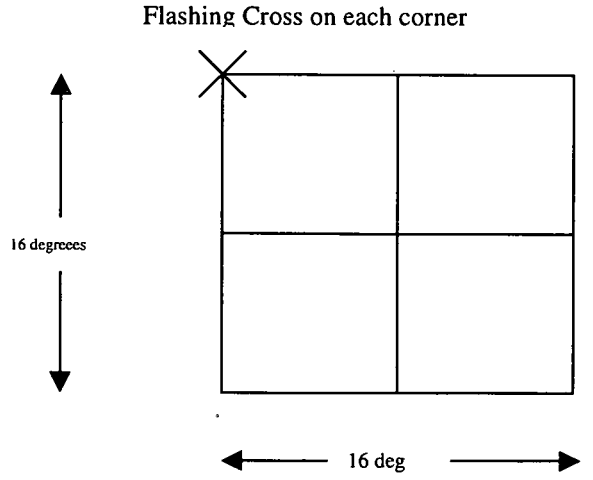


Figure C11 WF Evaluation Grid

To assist the calibration process a number of utility software programs have been written. The first of the utilities is a program called "Eye". This program is a menu driven program that runs in the control PC. The function of this program is to control the ESFID DSP during testing and calibration, and it is the software that interfaces with the DSP through the 8051 controller. The functions of "Eye" are shown on table C1.

MODE	FUNCTIONS	DESCRIPTIONS
DSP	Initialise Download FLASH	Initialise the DSP processor and 8051 processor Download the DSP code into FLASH memory
Calibration	Send NN Weights	Send the Neural Network weights to DSP RAM
Utilities	Set Mirrors Test Comms	Command the mirrors to user defined positions Test the RS232 interface to DSP
Eye-tracker	Run calibration	Run the eye-tracker calibration program described above

Table C1 "Eye" Functions

The next utility is the neural net training algorithm that is based on a Levenberg-Marcquard Back-propagation algorithm. This utility is called together with a error threshold value, which is a non-dimensional error value to which the training should converge. It was found that a good eye calibration data-set converged to within an error threshold of below 5 after about 10 iterations. It was also found that if the data-set took a long time (20 plus) iterations to converge, the actual eye-tracking performance was not to specification.

C.3 ESFID SIMULATOR

The ESFID Simulator block diagram is shown in figure 4.5. It is repeated here to show the elements. The simulator consists of the three PC's (the control PC, the wide-field simulation PC, the narrow-field simulation PC) and the RGB to TV converters that generate the video for the ESFID display. The simulator receives the eye angles over the RS422 link from the DSP, and the helmet angles from the RS232 link from the helmet tracker. The information is passed between the PC's through an ethernet network.

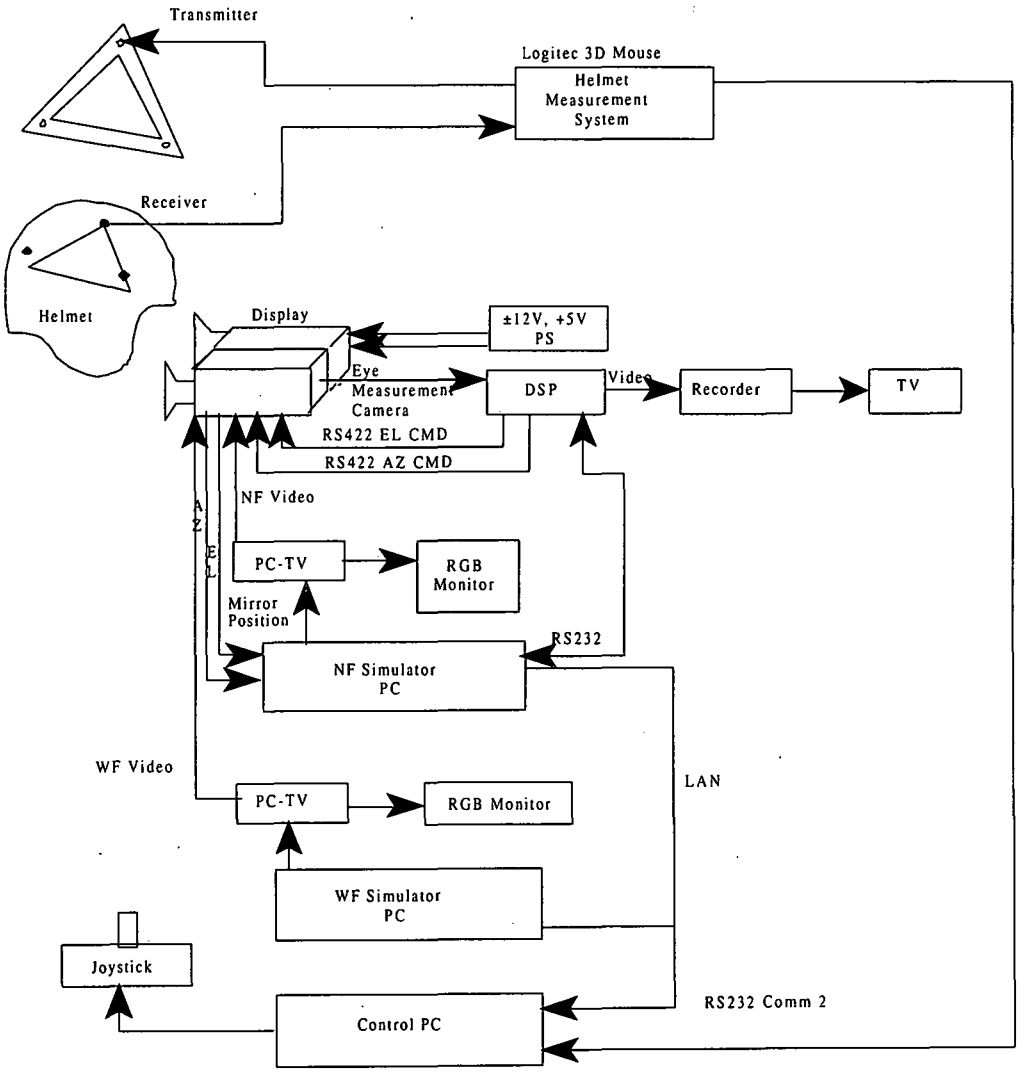


Figure C12 ESFID Simulator Block Diagram

The ESFID simulator was used to the target acquisition tests. The simulator was described in chapter 5.3. The purpose of this section is to describe the actual operation of the simulator for a target acquisition run. The software flow diagram is shown on figure C13.

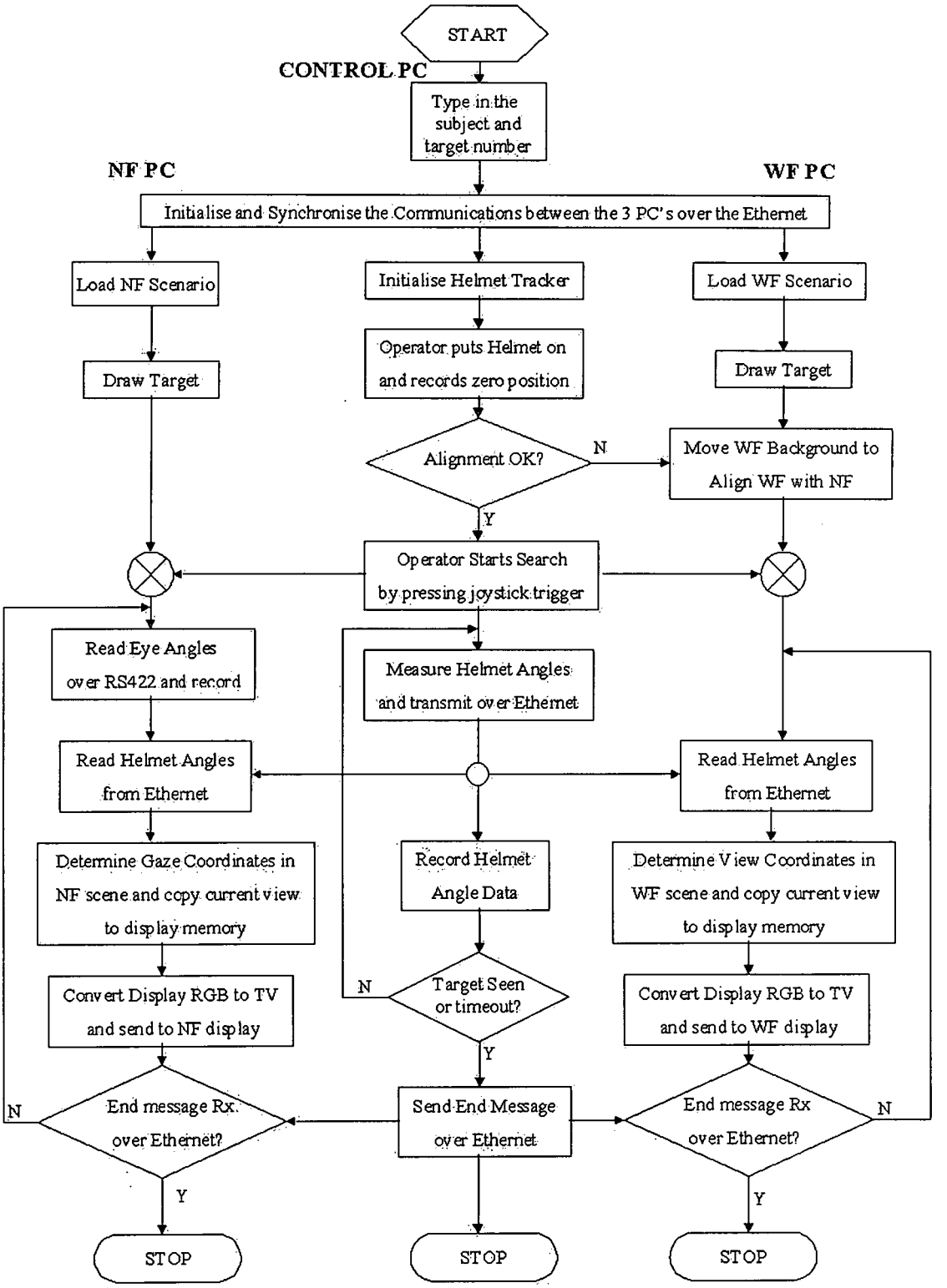


Figure C13 Simulator Flowchart

Powering up all the components in figure C12 starts the simulator. The control PC control program, called TELC is started and the operator enters the test subject's name and the target number. The next step is to initialise the ethernet network. The network contains the following information that the PC's use.

- Target number from the control PC
- Subject number
- Helmet AZ and EL angles from the control PC
- A MODE message indicating in which mode the simulator is in.

Once the network is up and running, the helmet tracker is initialised by the control PC, while the WF and NF PC's load their scenarios and draw their targets on the scenario at the correct position and range as indicated by the target number. The test subject is then asked by the simulator operator to place the ESFID helmet on his head and to move into a comfortable position. Moving the wide-field display with one-pixel offsets until the subject comfortably fuses the two displays electronically, aligns the wide-field and narrow-field displays. The process of alignment is done only once per session, and the alignment data is stored for that person, for later use. The alignment is also tested for robustness by moving the scenes out of alignment in every direction, until the test subject cannot fuse the pictures anymore. The alignment is then moved back to the centre of the range that the subject could hold fusion. The operator is then asked to zero the helmet measurement by pressing a button on the joystick. The actual search exercise commences on the next trigger from the joystick.

During search, the control computer reads the helmet positions and transmits it over the network for the other PC's. The delay is very small, less than 20ms. The NF computer reads the eye-tracker DSP eye angle measurements, and together with the helmet angle determines the gaze coordinates in the scenario. The appropriate view is then copied from the scene to the display memory of the PC. The wide-field computer does the same, except is uses the helmet angles to determine the gaze coordinates in the scene. The RGB outputs of the WF and NF simulator PC's are converted to video with RGB to TV converter units.

As soon as the subject acquires the target, the joystick button is pressed again, and an END mode command is transmitted over the network. The time for the search is recorded. If the time allocated for the search has run out, the same occurs and the search time is recorded as time-out.

The scene that is generated in the WF and NF PC's has been described in chapter 3.2.3. A target is superimposed on this scene using 3D computer graphics techniques, for both the wide-field and the narrow-field scenes. The target was generated from 322 polygons. The T72 tank target is shown in figure C14.

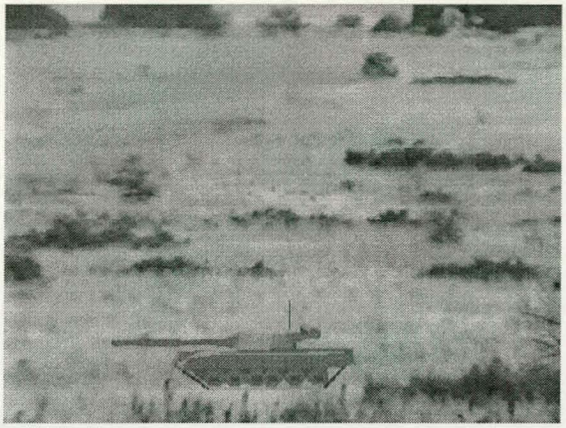


Figure C14 Graphics Target

The target/background inherent contrast was controlled by the average brightness (colour) of the target. The targets were also generated with the tank either facing to the right or the left, and the operator was asked to identify the target direction after he acquired the target, to ensure that the correct target was identified. The target was drawn on the scenario at its appropriate position and attitude. The lighting model used both the sun reflection and ambient light conditions on flat shaded polygons. The whole target was then blended into the background with a 2D anti-aliasing filter. The filter weighed each cell with 10% of the brightness of the cells surrounding the current cell.

The contrast between the target and the background was determined as the average brightness of the target divided by the average brightness of the background, which is of the same area as the target, around the target. This was accomplished by first drawing the target on a mask, to determine which pixels are target pixels and which are background pixels. The mask pixel coordinates are used to count the average brightness of the target and the background after the target has been drawn.