A System for Eye-Directed Control in a Split-Foveal-Peripheral-Display

by

Benjamin Nortjé

Submitted in partial fulfillment of the requirements for the degree

Master of Engineering (Electronic Engineering)

in the

Faculty of Engineering

UNIVERSITY OF PRETORIA

December 1999

Summary

TITLE: A System for Eye-Directed Control in a Split-Foveal-Peripheral-Display

AUTHOR: B Nortjé

SUPERVISOR: Prof JJ Krüger, Department of Electrical, Electronic and Computer Engineering, Faculty of Engineering

CO-SUPERVISOR: Prof PJ Cilliers, Department of Electrical, Electronic and Computer Engineering, Faculty of Engineering

DEGREE: Master of Engineering (Electronic Engineering)

UNIVERSITY: University of Pretoria, South Africa

ABSTRACT:

In this thesis an eye-directed controller is developed that slaves the narrow field display within a split-foveal-peripheral-display system to the operator's gaze position. A neural network controller is proposed that directly maps the gaze position to the narrow field projection co-ordinates without the need for any axis or co-ordinate transformations. A novel image feature-extraction algorithm, for extraction of the pupil-purkinje difference measure, has been developed that exhibits robust and reproducible real-time performance. By providing foveal and peripheral vision in a far-field teleoperator through the eye-directed split-foveal-peripheral-display, visual information is sufficiently and naturally provided for the establishment of telepresence.

Keywords:
Eye-directed control, split-foveal-peripheral-display, neural network controller, purkinje image extraction, eye movement, gaze determination, eye tracker, head-mounted display, foveal-peripheral fusion, telepresence, far-field teleoperator.

Opsomming

TITEL: 'n Oog-geslaafde Beheerstelsel vir 'n Verdeelde-foveale-perifere-vertooneenheid.

OUTEUR: B Nortjé

STUDIELEIER: Prof JJ Krüger, Departement Elektriese, Elektroniese en Rekenaar-Ingenieurswese, Fakulteit Ingenieurswese

MEDE-STUDIELEIER: Prof PJ Cilliers, Departement Elektriese, Elektroniese en Rekenaar Ingenieurswese, Fakulteit Ingenieurswese

GRAAD: Magister in Ingenieurswese (Elektroniese Ingenieurswese)

UNIVERSITEIT: Universiteit van Pretoria

ABSTRAK:

In die verhandeling word 'n oog-geslaafde beheerder ontwikkel wat die nouveld-vertoon in 'n verdeelde-foveale-perifere-vertoonstelsel slaaf aan die gebruiker se staar/kykposisie. Om die staar-posisie direk te koppel aan die nouveld-projeksie koördinaatstelsel word 'n neurale netwerkgebaseerde beheerder voorgestel wat die nodigheid vir enige koördinaatstelsel-transformasies totaal ophef. 'n Nuwe beeld-kenmerkonttrekkings-algoritme, met intydse robuuste en herhaalbaarheids-eienskappe, is ontwikkel om die pupil-purkinje kenmerk te onttrek. Met die daarstelling van beide foveale en perifere visie met die verdeelde-foveale-perifere-vertoon stelsel in 'n vêrsiende tele-operateur, word visuele informasie getrou en natuurlik oorgedra vir die daarstelling van tele-teenwoordigheid.

Sleutelwoorde:
Oog-geslaafde beheer, verdeelde-foveale-perifere-vertooneenheid, neurale netwerkbeheerder, purkinje beeld-onttrekking, oogbeweging, staar/kyk-bepaling, oogvolger, kop-gemonteerde vertooneenheid, foveale-perifere fusie, tele-teenwoordigheid, vêrsiende tele-operateur.

Acknowledgements

I would like to thank the following persons and institutions for their invaluable help in making this study a success.

1. Prof Johan J Krüger for his vision, patience and leadership.
2. Prof Pierre J Cilliers for his motivation and support through all the years.
3. Gerrit T Viljoen for his advice, support and who made this work possible.
4. Kentron for their financial support during the first two years.
5. Alan Kuyper for his continued support.
6. Suleen for your extreme patience and support.

Contents

Contents

Contents

List of Abbreviations

CCD	charged coupled device
CRT	cathode ray tube
DSP	digital signal processing
FOV	field-of-view
IR	infrared
LED	light emitting diode
NF	narrow field
NN	neural network
PC	personal computer
PPD	pupil-purkinje-difference
RMS	root mean square
SFPD	split-foveal-peripheral-display
VFG	video frame grabber
WF	wide field

1.1 Introduction

In a remotely operated system visual communication is indispensable and needs to be presented as naturally as possible to the operator. This is often defined as telepresence, which Sheridan [1] describes as when: "The human operator receives sufficient information about the teleoperator and the task environment, displayed in a sufficiently natural way, that the operator feels physically present at the remote site." A teleoperator is defined as a machine that operates on its environment and is controlled by a human at a distance [1].

In recent years various head-mounted displays have been developed in order to provide an enhanced visual man-machine interface. In our research we want to develop and refine a far-field teleoperator that can typically be used as an observation sight or as an aid in remotely piloted vehicle control. In these circumstances the visual object distances usually exceed 100 m where vision becomes monoscopic [2], implying that a monocular display system can be used within the far-field teleoperator. In near-field tele-existence systems a binocular visual space is a prerequisite [3, 4].

In current far-field teleoperator systems monoscopic vision is provided via wide field-of-view (FOV) optics. The typical elements for such a system is depicted in figure 1.1.

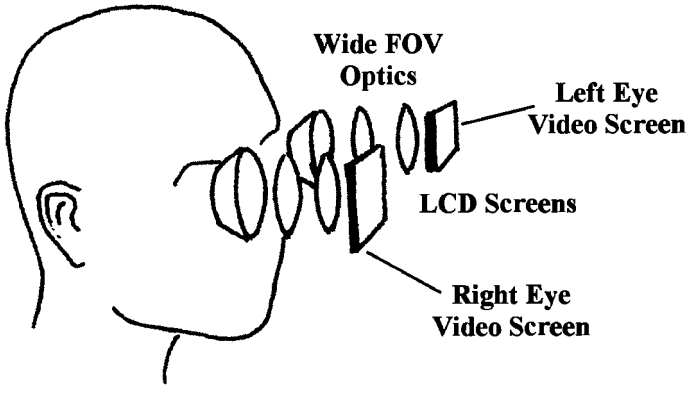

Figure 1.1 Elements of a typical head-mounted display. [5]

1.2 Objective

In order to induce telepresence as naturally as possible in a far-field teleoperator we must provide the operator with both foveal and peripheral vision. This can be accomplished with current technology via the SFPD concept.

The objective of this dissertation is to present the design, implementation and test of an eye-slaved controller for a SFPD system. A system must be developed that slaves a high-resolution foveal image to a subject's gaze position within a wide field scenario. The SFPD system consists of a helmet head-up display, where the wide field scenario image is presented to the subject's left eye and the narrow field image to his right eye. The wide field or peripheral display is updated with head movement, while the narrow field display is imposed on the peripheral image at the gaze position. The SFPD system used in this dissertation has been designed by Viljoen and is presented in [10]. It was my duty and privilege to design and integrate an eye-directed controller for the system. In short, the eye-directed controller has to determine the point of gaze of the operator in the SFPD, in order to provide the control commands that are used to position the narrow field display and update the displayed image.

The control commands consist of 12-bit, serial co-ordinates that project/position the narrow field display at the gaze position. The main assignments in this dissertation can be summarised as follows:

- Implement a robust and real time feature extraction algorithm to determine the subject's gaze position within the display.
- Design and implement a control system with the gaze position as input and the projection control commands as output.

In the first phase of this project a prototype eye-tracker device was built to get acquainted with this field and to gain experience in gaze measurement. This initial work provided the stepping stone for the main design in the SFPD system. In this design process a new image feature extraction algorithm for gaze measurement has been developed that enables robust

real-time eye feature extraction. A neural network controller was proposed and implemented for mapping the gaze position into real-world camera co-ordinates.

As stated above, gaze determination forms an integral part of this dissertation as the person's gaze position within the image displays serves as the system's input. In order to accomplish this a basic knowledge of the physiology of the eye and eye movements is required. Chapter two provides the necessary background on these topics. A survey of eye measurement techniques used for gaze determination is given in chapter three, as well as a description of the preliminary design and verification process followed in the prototype eye tracker design.

In chapter four the SFPD concept is described in more detail. The SFPD experimental setup is discussed and the final system is shown. The conceptual design and methodology followed in the dissertation is described in chapter five. This includes a description of the image feature extraction algorithm development used in determining the gaze position and a discussion of the neural network implementation used to map the gaze position to the SFPD co-ordinates. The evaluation procedure and results obtained in the final tests are presented in chapter six; followed by the discussion and conclusion in chapter seven.

An example of a possible far-field teleoperator scenario is depicted in figure 1.2. In this example a wide field (WF) and narrow field (NF) camera is attached on top of the vehicle. The NF camera can be manoeuvred within the wide FOV and is slaved to the operator's gaze position. Through this configuration telepresence can be established, because visually the operator receives the same information, as he would have being present physically.

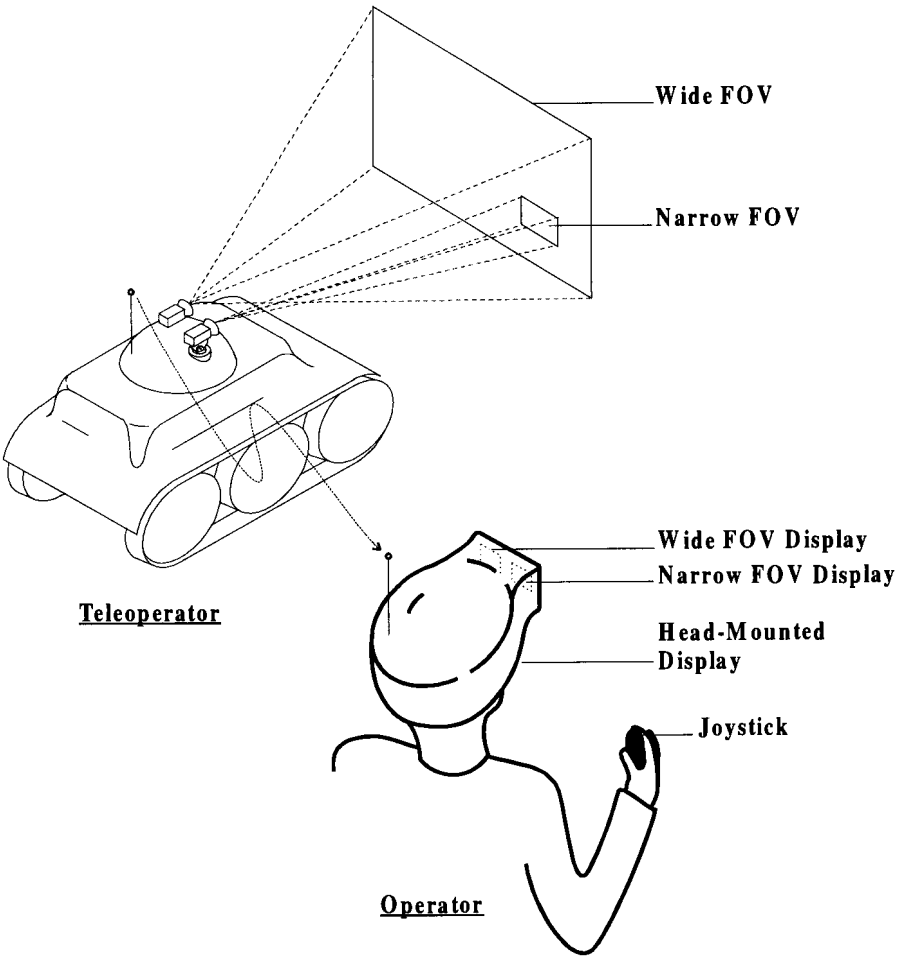

Figure 1.2 Far-field teleoperator implementation with both wide and narrow FOV capability.

Chapter 2

Physiological Background

"The eye has often been compared to a camera. It would be more appropriate to compare it to a TV camera attached to an automatically tracking tripod - a machine that is self-focusing, adjusts automatically for light intensity, has a self-cleaning lens, and feeds into a computer with parallel-processing capabilities so advanced that engineers are only just starting to consider similar strategies for the hardware they design."

D.H. Hubel [11]

The first step in any design process starts with a system specification process. In this process one needs to describe the system's behavior in order to understand what to design and implement. Clearly, if one wants to determine gaze direction, one requires a sound understanding of the eye anatomy and physiology.

In this chapter we take a brief look at the eye structure and the characteristics of eye movements. A substantial part of the anatomy and physiology section has been obtained from Marieb [12], while the eye movement section mainly comes from Young and Sheena [13].

2.1 The Eye and Vision

Seventy percent of all the sensory receptors in the body are in the eyes. Photoreceptors in the eye transforms and encode incoming light, from where the information is carried on massive optic tracks towards the brain. The surface anatomy of the eye is shown in figure 2.1, with indications to the relevant accessory structures of the eye. The eyebrows shade and protect the eyes. Anteriorly, the eyelids protect the eyes and reflex blinking of the eyelids helps prevent the eyes from drying out. Eyelashes on the border of the eyelids trigger reflex blinking when anything touches them.

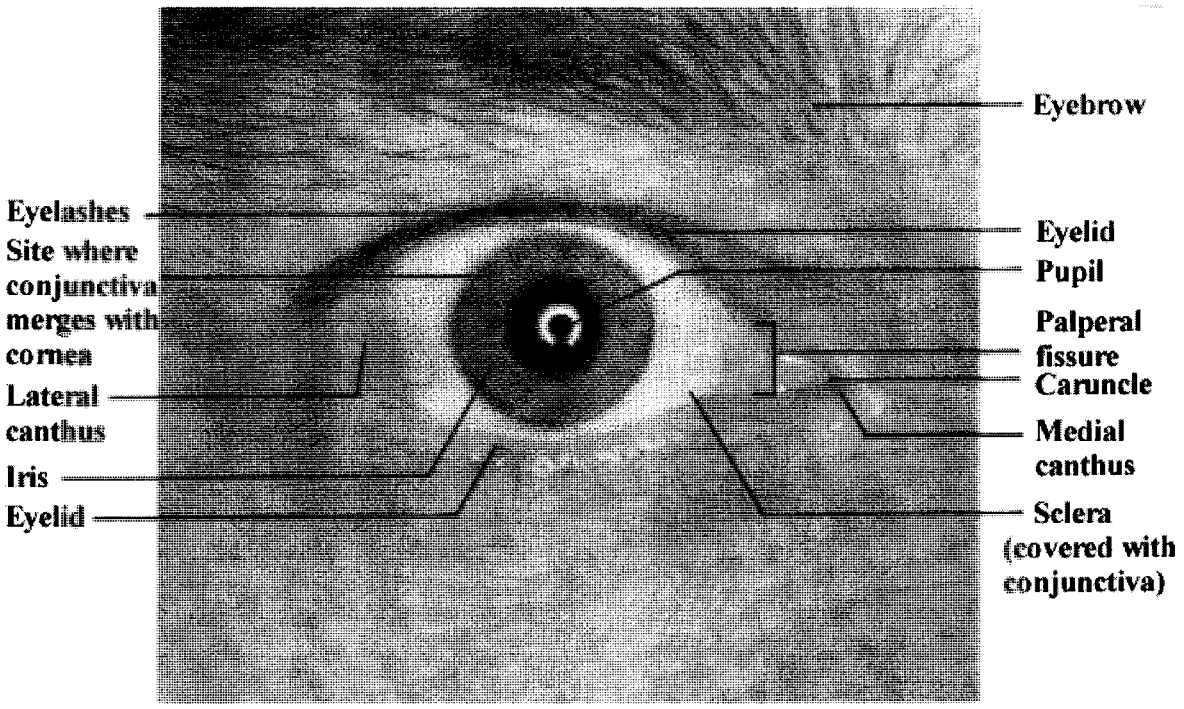

Figure 2.1 Surface anatomy of the eye. [12]

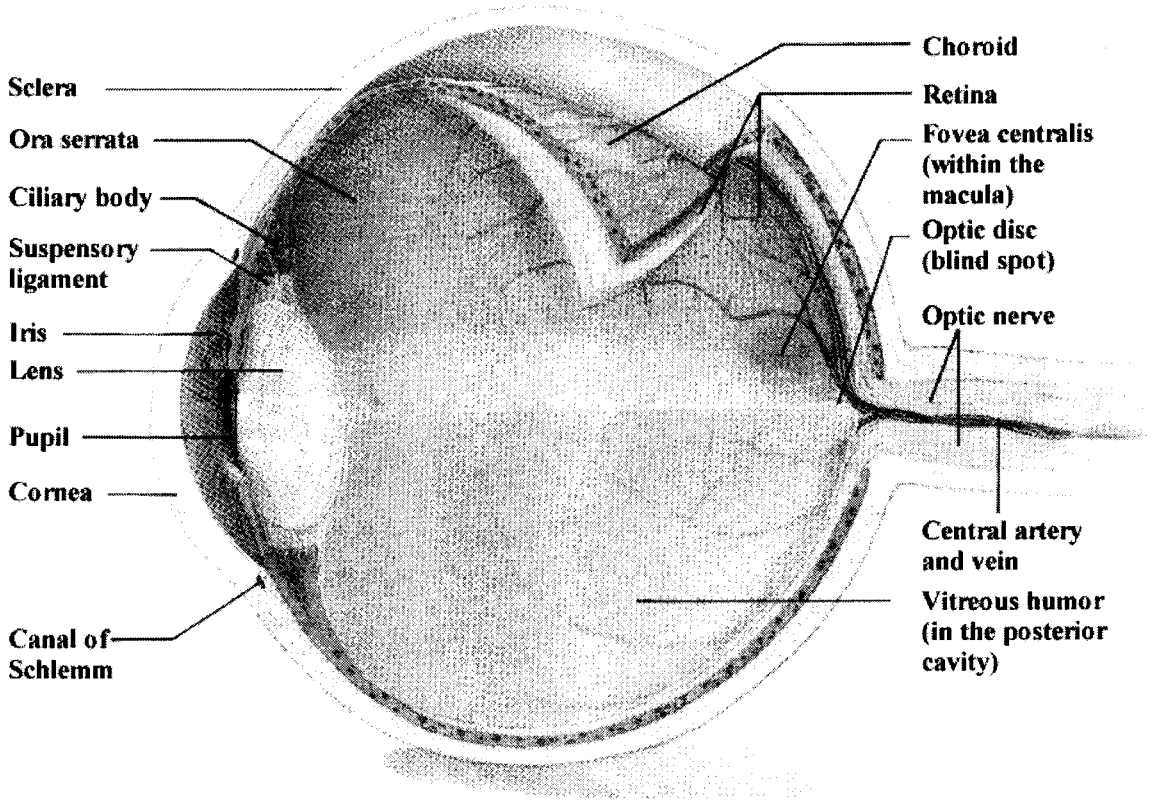

Figure 2.2 Internal structure of the eye. [12]

2.1.1 Eye structure

The entire eye is a spherical structure approximately 2.5 cm in diameter. *(See figure 2.2.)* The outer part of the eye is composed of three layers of tissue. The sclera is the protective outside layer and covers about five-sixths of the surface of the eye. Seen anteriorly as the "white of the eye," the sclera protects and shapes the eyeball and provides a sturdy site for anchoring the extrinsic eye muscles. Anteriorly the sclera forms the transparent cornea that allows light to enter the eye. Posteriorly the sclera is pierced by the optic nerve.

The middle layer of the coating of the eye is the uvea and has three distinct regions: the choroid, the ciliary body and the iris. The choroid is a deeply pigmented membrane and is located at the posterior part of the eyeball. It helps to absorb light, preventing it from scattering within the eye or leaving the eye. Anteriorly it is continuous with the ciliary body that encircles and controls the lens. The lens itself is a flattened sphere constructed of a large number of transparent fibers arranged in layers. The lens is a biconvex flexible structure that can change shape to allow precise focusing of light on the retina.

The iris is the most anterior part of the uvea and lies between the cornea and the lens. This visible colored part of the eye has a round central opening, the pupil, that allows light to enter the eye. The iris is made up of circularly and radially arranged smooth muscle fibers and acts as a reflexively activated diaphragm to vary pupil size. In bright light and close vision conditions these muscles contract and the pupil constricts. Vice versa in dim light and distant vision conditions, the pupil dilates and allows more light to enter the eye. Normally pupillary action takes place nearly equally in the two eyes, even if only one eye is subjected to illumination changes.

Reflex changes in pupil size may also reflect emotional or interest reactions to what is viewed. When the subject matter is appealing or during problem solving, pupillary dilation frequently occurs. Although irises come in different colors, they contain only brown pigment. Eyes appear brown or black when they contain a lot of pigment. With less pigment present the unpigmented parts of the iris scatter the incoming light and eyes can appear blue, green or gray.

The innermost layer is the light-sensitive retina, which in itself has a inner neural layer and an outer pigmented layer that absorbs light. The neural layer is composed of three main types of neurons: photoreceptors, bipolar cells, and ganglion cells. Incoming light that is passed through the pigmented layers induces currents in the photoreceptors, which are passed on to the bipolar neurons and then to the innermost ganglion cells. The action potentials from the ganglion cells leave the eye via the optic nerve that exits the eye at the posterior wall of the eye. This exit spot is called the optic disc or blind spot and does not contain any photoreceptors; therefore light focused on the optic disc cannot be seen.

Two types of photoreceptors are found in the neural retina: rods and cones. Cones operate in bright light and provide high acuity color vision. From each cone there is a single fiber-pathway which, passing through several synapses, retains its identity in the optic nerve to the brain. This fact permits fine differentiation of spatially separated stimuli that fall on cone areas. The more numerous rods have a high sensitivity to light and an appreciation of moving objects [13] that allow us to see rather fuzzily in gray tones. Peripheral vision is rod vision.

Lateral to the blind spot is the macula lutea region that contains the fovea centralis in its center. The macula mostly contains cones, while the fovea only contains cones. Cone density rapidly declines from the macula edge toward the retina periphery and vice versa for the rod density. The retina periphery contains only rods. Anything that we wish to view critically must therefore be focussed on the fovea. The fovea is quite small and subtends a visual angle of about one-degree. As each fovea is only about the size of the head of a pin, not more than a thousandth of the entire visual field is in hard focus at a given moment. In order to visually comprehend a rapidly changing scene, the eyes must flick rapidly back and forth to provide the foveae with detailed images of different parts of the visual field.

2.1.2 Eye muscles

Each eye has six extrinsic eye muscles attached to it. These muscles make it possible for the eye to follow a moving object. The extrinsic eye muscles are among the most precisely and rapidly controlled skeletal muscles in the entire body. In order to track an object a precision of a few minutes of arc is required, or else double vision would result. Such precise

movements require a collection of finely tuned reflexes, including those that control head position. [11]

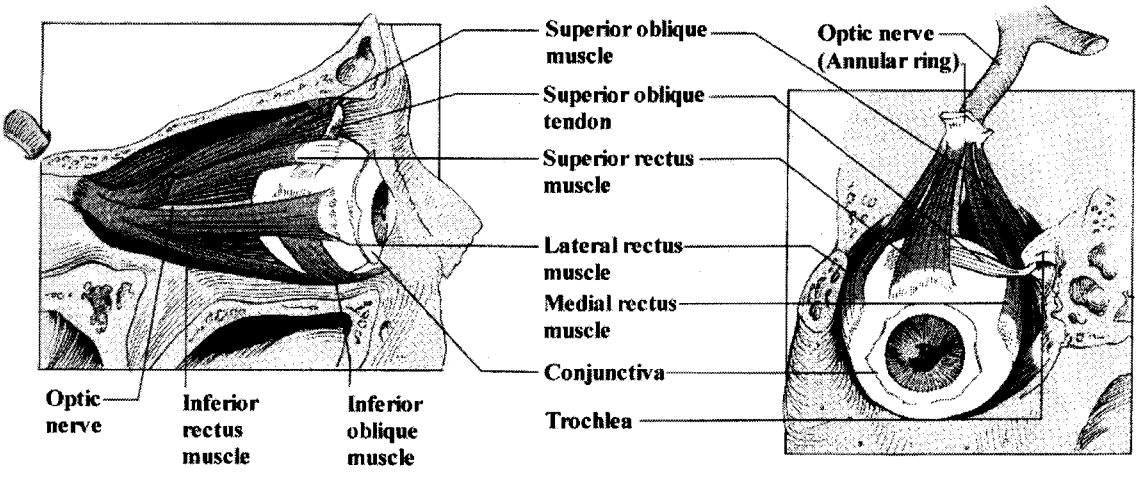

Figure 2.3 Extrinsic muscles of the eye. [12]

Eye movement is performed by contraction of one muscle and relaxing its opponent by the same amount. Each eye muscle is made to contract by the firing of motor neurons in the brainstem. The muscles take care of movements in one of three orthogonal planes and the movements that they promote are indicated by their names: superior, inferior, lateral and medial rectus muscles, as well as two oblique muscles. The medial-lateral rectus muscle pair controls horizontal eye movements; the superior-inferior rectus muscle pair controls vertical eye movement; and the oblique muscles control torsion about the visual axis. As the muscles pairs are not arranged orthogonally, coupling of movement occurs between planes. [15]

2.1.3 The visual pathway

The optic nerves which exit from the eyeballs merge to form the optic chiasma, as shown in figure 2.4. At the optic chiasma the fibers from the medial aspect of each eye cross over and continue via the optic tracks to the thalamus. Each optic track therefore carries medial information from one eye and lateral information from the other eye. The left optic track thus carries a complete representation of the right half of the visual field, and vice versa for the right optic track. Via synapses the information is passed on to the thalamus in the lateral geniculate body, from where neurons carry it to the occipital lobes of the cerebrum.

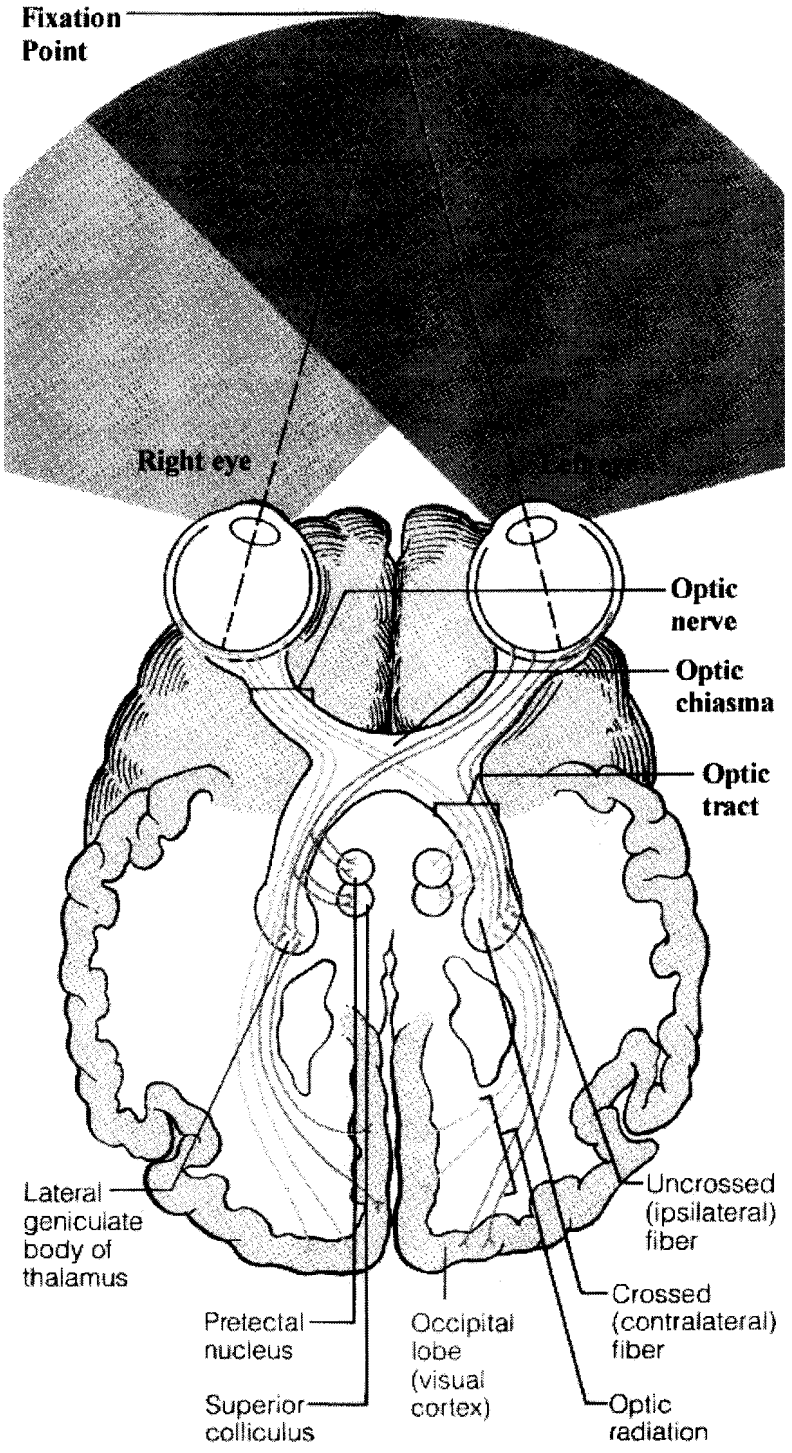

Figure 2.4 Visual fields of the eyes and the visual pathway to the brain. [12]

Many nerve fibers directly enter the midbrain and terminate at the superior colliculi and pretectal nuclei, which respectively direct the visual reflex control of the extrinsic muscles of the eyes and mediate the pupillary light reflexes. Notice that the visual fields of the eyes overlap considerably and that each eye sees a slightly different view (area of binocular vision). As each visual cortex receive information from both eyes, depth perception and accurate distance judging are possible. This faculty, known as stereopsis, is a result of cortical fusion of the slightly different images delivered by the two eyes and depends on the two eyes working together and accurately focusing on the object.

Thalamic processing at the lateral geniculate bodies mostly handles depth perception and contrast enhancement activities. There are many more synapse connections with the center of the retina, than those from the retina periphery. It therefore appears that the thalamic nuclei are mostly concerned with high-acuity aspects of vision. Cortical processing occurs in two types of cortical neurons: simple cortical neurons responds to stimuli like straight edges of light and dark contrast conditions, while complex cortical neurons responds to higher abstract stimuli like the motion of objects in the visual field. Figure 2.5 depicts the different stages in the visual pathway.

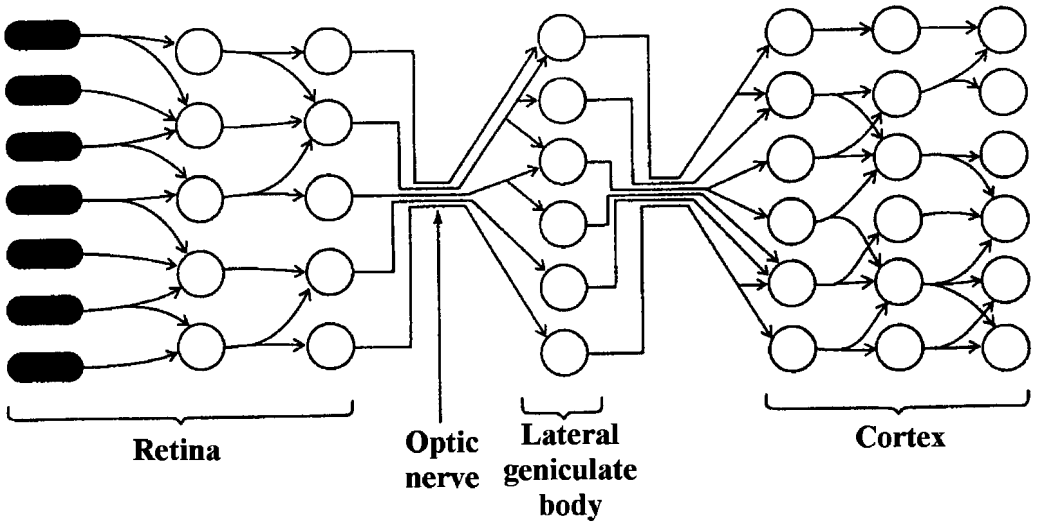

Figure 2.5 Stages in the visual pathway. [12]

2.2 Eye Movements

Most cortical cells respond better to movement than to a stationary object, as movement in the outside world is much more important than a static condition. With so many oriented cells insensitive to static input, it is quite interesting to see how we do view a stationary scene.

Instead of examining a scene with smooth continuous movements, the eyes fixate on an object so that the image stabilizes on the foveae. Within a brief period the eyes jump to a new position and fixate on that point. The new position could have been triggered by slight movement, an interesting shape or by contrasting with the background. During the jump or saccade the eyes move so rapidly that our visual system does not even respond to the resulting movement of the scene across the retina [11]; we are altogether unaware of the violent change. A visual scene is therefore examined by jumping and fixating from one place to another. These fixations or search patterns are also known as scan paths, and various studies have been conducted in the analysis, prediction and efficiency of individual's search patterns. [6], [16], [17], [18]

The center picture in figure 2.6 shows the monitored eye positions of an observer on viewing the picture shown on the left. On the right the eye movements of an observer are shown for the first time viewing of the given drawing.

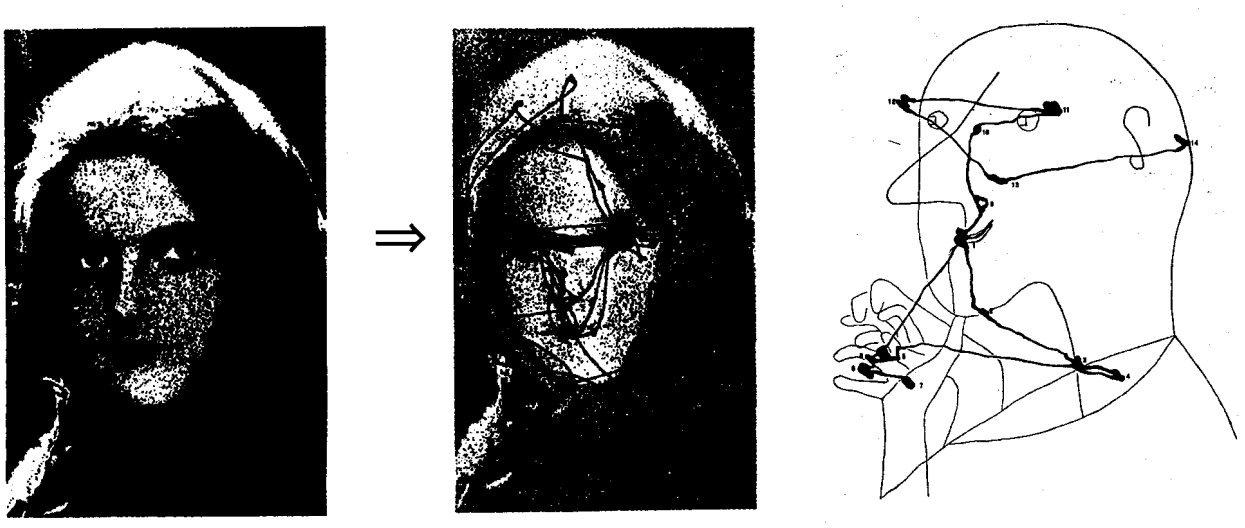

Figure 2.6 Monitored eye movements on viewing picture. [11], [18]

Despite the efforts we may make, our eyes do not hold perfectly still during fixation on a point of interest in a stationary scene. Constant tiny eye movements or microsaccades occur several times per second and are more or less random in direction [11]. In experiments during the fifties it was shown that when any movement relative to the retina is eliminated by means of artificial stabilization, vision fades away and the scene becomes black. Thus, however counterintuitive it may seem, microsaccades are necessary for viewing stationary objects.

Eye movements are categorized according to their motion response and the stimuli that provoked the response. Four systems are responsible for generating the types of eye movements seen in a normal individual: [15]

- Vestibular system; sensitive to head movement in space.
- Optokinetic reflex system; responds to movement of a visual field across the retina.
- Pursuit system; responds during fixation of a target.
- Saccadic system; responds during fixation of a target.

In the following paragraphs a very brief description of only the main eye movement types is given, as these are the typical movements encountered in our system.

Saccadic eye movements are characterized by very high initial acceleration and deceleration endings (up to 40 000 deg/sec^2). Peak velocity during saccades may attain 400-600 deg/sec, while the duration of a typical saccade may vary from 30 to 120 msec. During searching saccadic movements are in the range of 1-40 deg; head movement is often involved when target motion exceeds 30 deg.

It is very interesting that in response to a visual stimulus, saccadic eye movement exhibit a latency of 100-300 msec and that the minimum delay or refractory period between saccadic movements vary between 100-200 msec. Typically there are two or three saccades per second, but they only occupy 10 percent of the viewing time because of their high-speed [18].

Since the faculty of vision is greatly impaired during saccades (a phenomenon known as saccadic suppression), the majority of visual inspection occurs during fixations [7]. In [21] the question is asked why do saccades take such a long time; if saccades have this tremendous

speed and accuracy, why does it take such an extraordinary long time to react to a stimulus? This latency effect is also known as oculomotor procrastination.

"A typical saccade of 10° lasts about 50 msec, during which the high rate of visual slip, amongst other factors, renders the visual system all but blind. Thus, paradoxically, the more frequently the oculomotor system tries to improve things by getting the fovea exactly on target, the less time remains actually to see: so that a suitable balance must be struck between not seeing quite what we want, and not seeing at all. " [21]

Pursuit movements, in the range of 1-30 deg/sec, are used to track slowly moving targets. These are smooth movements that stabilize the image of the moving object or background on the retina, independent of the saccadic system. These movements usually require a moving stimulus for execution and generally are not under voluntary control. Closely related *compensatory eye movements* in turn tend to stabilize the retinal image of fixed objects during head motion and react on semicircular-canal stimulation.

Vergence eye movements occur when the eyes move in opposite directions for image fusion on near or far objects and have a maximum velocity in the order of 10 deg/sec over a 15 deg range.

Miniature eye movements or fixation movements occur during attempted steady fixation on a target and are generally less than 1 deg in amplitude. *Microsaccades* redirect the eye toward the necessary position for fixation on the target, with minimum separation periods of 30 msec. Microsaccades can have magnitudes as large as 1 deg. Because of the presence of these fixation movements, an accuracy of 0.5 to 1 deg is often sufficient in eye-monitoring tasks designed to show what part of the visual field is being fixated [13].

Optokinetic nystagmus is evoked when a moving visual field contains repeated patterns. (For example when one looks out of a train window and maintains fixation on an object.) These movements have a slow phase in which pursuit movement is used for fixation, and a fast return phase in which the eye exerts a saccadic jump to fixate on a new portion of the scene. The maximum frequency between the fast phases is 5 Hz.

Eye movements that have similar amplitude and frequency characteristics as optokinetic nystagmus are *Vestibular nystagmus*. Vestibular nystagmus is elicited with deflection of the capulas of the horizontal semicircular canals with head rotation. Counterclockwise head rotation about the vertical axis will cause clockwise slow-phase eye movement. Return saccades will occur if the head motion continues.

Torsional eye movements are rotation movements of the eye about the line of gaze and may be stimulated by rotational optokinetic nystagmus or vestibular responses.

Chapter 3

Eye Tracker: Preliminary Design

In the previous chapter a brief overview has been given on eye movement characteristics and on how we observe a visual scene. Eye movements have been researched in various fields, ranging from psychological analysis to being applied in computer interface aids for the handicapped.

> "*The main reason for eye-gaze based user interfaces being attractive is that the direction of the eye-gaze can express the interest of the user – it is a potential porthole into the current cognitive processes – and communication through the direction of the eyes is faster than any other mode of human communication. It is argued that eye-gaze tracking data is best used in multimodal interfaces where the user interacts with the data instead of the interface, in so- called non-command user interfaces.*" [27]

There are several techniques available for measuring eye movements or gaze direction. These techniques vary in accuracy, resolution, real-time response and ease of use. The ideal eye tracker device has a long list of usability requirements [27]. One particular method can't possibly satisfy all criteria and tradeoffs must be made for a specific application. In the SFPD system the main criteria for the eye measuring technique are:

- Minimum discomfort. *(As natural as possible.)*
- Real time response integration.
- Root mean square (RMS) accuracy of 1°.
- Repeatability and robustness.
- Integration within head-up display environment.

In the first section of this chapter a survey is given on various eye measurement techniques. This is followed by the description of the preliminary design and prototype apparatus used in the eye tracker development.

3.1 Eye Measurement Techniques

Most of the information in this section has been obtained from Young and Sheena's work in [13].

3.1.1 Electro-oculography

In electro-oculography (EOG) techniques the corneo-retinal potential is measured to determine the eye position. The potential or electrostatic field that rotates with respect to the eye is measured by placing skin electrodes around the eye, through which the potential differences can be recorded. An electrode pair is required for each axis, therefore two orthogonal placed pairs are required for horizontal and vertical eye movement detection. The potentials are in the range of 15 to 200 mV.

Normally EOG implicates a dc recording method in order to determine eye position, whereas in electronystagmography (ENG), ac measurements are used for detailed eye movement recordings. EOG recordings have several undesirable factors that hamper recordings:

- Large muscle-action-potential artifacts are picked up by the skin electrodes and sometimes make it difficult to detect the EOG signals. (Proper electrode placement can reduce muscle activity artifacts.)
- External electrical interference are troublesome, unless the system is properly shielded.
- When both horizontal and vertical eye movements are recorded, errors are caused by axis coupling and by the nonlinearity of the records.
- EOG measurements are corrupted by base-line drift.
- Motion of the upper eyelid causes overshoot artifacts in vertical recordings.
- The corneo-retinal potential varies with changes in the environmental lightning.

Without requiring visualization of the eye, like in most other techniques, EOG can measure eye movements up to ±70 deg and has a typical accuracy of ±1.5-2 deg. Clinicians prefer EOG recordings for their needs, as it is a simple, low cost technique that does not require any head monitoring equipment [15]. This technique is rather troublesome and is not well suited for everyday use since it requires the close contact of electrodes to the user [27].

3.1.2 Head-mounted or Head-fixed Reflected Light Techniques

In these techniques a light source is fixed with respect to the subject's head in order to visualize specific regions in the eye and either require a bite board or head strap device.

3.1.2.1 Limbus Tracking

The limbus is the sharp boundary between the iris and sclera which can easily be optically detected and tracked. Usually the eye is illuminated with invisible infrared (IR) light. All limbus trackers measure the position of the limbus relative to the photodetectors or illuminators and the measurement is of the eye relative to the head. The main problem or obstruction in limbus tracking is the fact that the upper and lower boundaries of the iris are usually occluded by the eyelids and thus only provide precision tracking of horizontal movements.

Often eyelid position detection is combined with limbus tracking to aid in vertical eye movement measurement. In [28] however, significant disparity is reported between the eyelid and the vertical eye position. The success of this technique relies heavily on the configuration and use of the photoelectric elements. Often the photoreceptors are mounted on spectacle frames and require that the subject must relax the muscles around the eyes. Contradictory performance indications have been reported on this technique, with accuracy varying between 0.5-7° and a range of ±15°.

3.1.2.2 Pupil Tracking

The pupil tracking method is very similar to the limbus method and uses the boundary between the pupil and iris to detect eye gaze position. The center of the pupil virtually coincides with the foveal optical axis of the eye [13]. The 5-6° variation between the two can be corrected via proper calibration.

Under IR lightning conditions the boundary between the pupil and iris is sharper than that between the pupil and iris. Another distinct advantage over limbus tracking is that the range for eye movement detection is much greater because the pupil is smaller and eyelid occlusion occurs more infrequently. In some eye trackers collimated illumination is used. Under these conditions light is reflected from the interior of the eye and the pupil appears very bright; akin

to the "red eye" effect often seen in flash photographs where the flash lamp is close to the camera lens.

As mentioned in chapter 2, the pupil diameter varies as a result of both psychological and physiological influences and can hamper the calculations of the pupil centroid. A range of ±30° is possible for this method, with an accuracy of less than 0.5° mentioned in [27].

3.1.2.3 Corneal Reflection or Purkinje Image Tracking

In the limbus and pupil tracking techniques it is important that the head position must be fixed with regard to the measurement equipment. If the device is e.g. attached to a helmet even the slightest helmet slip can result in large errors. It also implicates that the system must be recalibrated each time the subject uses it. In most instances the gaze direction within the visual domain must be known, rather than the position of the eye with respect to the head. Clearly this is a function of both head and eye rotation; thus gaze direction can be deduced by summing the relative position of the eye to the position of the head.

"One of the main problems of measuring eye direction optically is that of separating lateral motion of the eye relative to the observer or the sensor and rotary motion of the eye relative to the scene" [13]. Therefore a technique is required that only measures eye rotation and that is not sensitive to head translation.

A technique that fulfills this requirement is to measure the difference between the centroid of the pupil and the corneal reflection (1st Purkinje image). Light that enters the eye is reflected from the front and rear surfaces of the cornea and lens, as shown in figure 3.1. These reflections are called *Purkinje images*.

In some eye tracking methods only the first Purkinje image is tracked. These methods exhibit results similar to the pupil and limbus trackers. One of the main problems in this technique is to obtain a good view of the eye. It also suffers from the same restrictions imposed on in the pupil tracking method in determining the centroid of the pupil. An accuracy of 1° is possible with a range covering of ±30°.

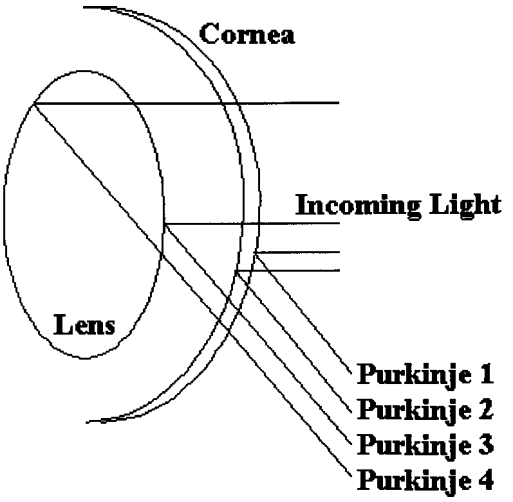

Figure 3.1 Purkinje Image Reflections

The gaze direction can also be calculated by using the relative positions between the first and fourth Purkinje images. This technique is more accurate than the other methods with a resolution of 1 min of arc and a range of ±20°. Unfortunately the fourth Purkinje image is very weak and requires special setup conditions. In some cases topical drugs must be used to dilate the pupils in order to expose it. The effects of eye translation and rotation are shown in figure 3.2. In the top row both eyes look straight ahead, with the right eye laterally displaced. In both eyes the Purkinje image is in the center of the pupil, indicating head translation independence. In the bottom row the displaced Purkinje image is shown with eye rotation.

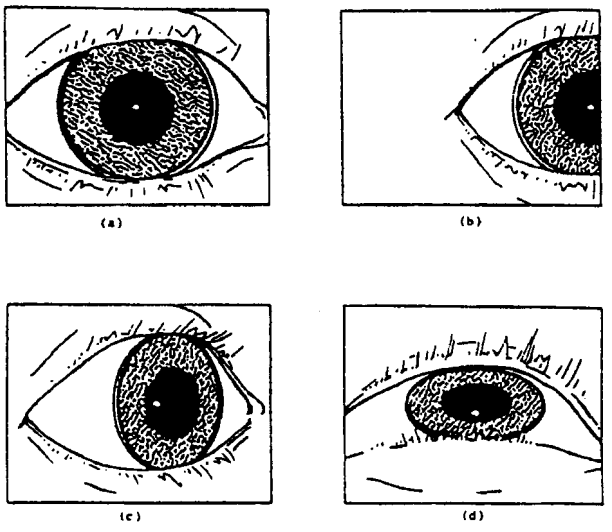

Figure 3.2 Effects of eye translation and rotation on the Purkinje Image [13]

u15761411
u15212646

3.1.3 Contact Lens Methods

In these techniques the subject wears special contact lenses. In one instance a magnetic search coil is attached to the lens and emerged in an alternating magnetic field that produces a voltage proportional to the sine of the angle between its plane and the field direction. In another method one or more mirror planes are engraved on the contact lens and reflections of light beams can be used to calculate the eye position.

Although quite accurate recordings are possible, these techniques are uncomfortable and certainly too cumbersome for everyday use. A resolution of 0.25° is possible with a range of ±50°.

3.2 Eye Tracker Prototype Development

In the SFPD system any of the selected users must be able to put on the helmet display and operate the system interchangeably; implicating that the measurement technique must have the minimum physical contact. This restricts the useable techniques to the head-mounted techniques discussed in the previous section.

The decision fell on the well-known corneal reflection and pupil relationship technique [13], [15], [29], [27]. In this technique the eye is illuminated with infrared light and the direction of gaze can be calculated from the difference between the centroid of the pupil and the corneal reflection or first Purkinje image. As described, the pupil-purkinje-difference (PPD) between the pupil centroid and first purkinje image centroid provides us with gaze positions invariant to small head translations.

The PPD phenomena can see easily be observed at night. Look at a person's eyes and let him fixate on various objects in a room in the direction of the light fixture. Notice how the corneal reflection position changes with respect to the pupil area. Gaze position can be directly linked to the eye position, or in our case to the PPD. In the proposed system the eye position of a subject must be mapped to the subject's gaze position within a distant scene. Before the design and manufacturing of the SFPD head-mounted system started, we constructed a prototype platform to evaluate and verify the eye tracker's operation.

A CCD (*charged coupled device*) camera is used to extract the relevant eye features needed to determine the gaze position. In view of the real-time constraint a dedicated video frame grabber (VFG) board was developed to fulfil the digital signal processing (DSP) requirements. A 60MHz TMSC31 processor is used, along with the necessary peripheral devices for servo control and communication that would be required in the final system.

3.2.1 Prototype Platform Configuration

A schematic representation of the prototype apparatus is given in figure 3.3. An infrared (IR) LED is used to illuminate the eye. The subject is positioned in front of a computer monitor and views the display through a transparent dichroic mirror. The mirror is coated with a layer that reflects infrared light and passes light in the visual spectrum. The illuminated eye's image is reflected at the beam splitter and provides an overhead camera with a view of the eye. The LED is positioned alongside the small CCD camera, thus forming a bright first-purkinje image to the right of the pupil centre. Head movement is inhibited via a chin rest and forehead fixture.

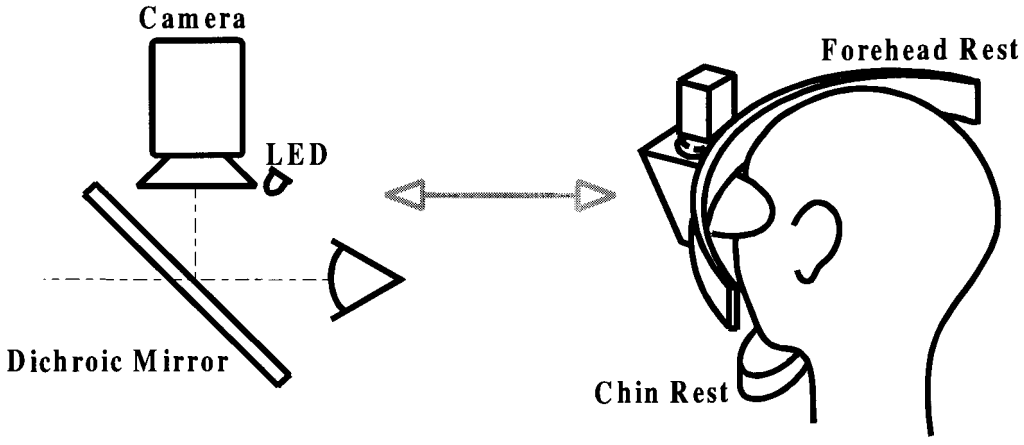

Figure 3.3 Prototype apparatus

Under IR light conditions nearly all the light is absorbed within the eye and the pupil appears much darker than under normal, visible light illumination; as described in [13]. This enhances the contrast between the pupil and iris areas, even in instances where a person has very dark irises. In normal light conditions it is often difficult to distinguish the pupil/iris boundary in persons with very dark irises, but under IR light the areas are clearly distinguishable.

The VFG board captures even and odd video frames with a resolution of 512x256 pixels. (*50 Hz interlaced update rate*). Therefore a 20 ms time frame exists in which all processing must be performed. The VFG board provides a video output channel, which enables the operator to view images of subject's eye, as well as any changes made to the frames during the processing stage. A captured frame of an operator's eye is shown in figure 3.4.

Figure 3.4 Extracted Eye Image Frame with IR Illumination

The purkinje image, the bright spot on the right, is quite distinct. The PPD is simply the vector difference between the centroid of the purkinje spot and the centroid of the dark pupil area.

3.2.2 Initial Prototype Algorithms and Results

The first step in the verification process was to develop a pupil centroid tracker, which imposed a crosshair at the centroid co-ordinate on the video output for evaluation purposes. (*Please note that the relevant algorithms are discussed in chapter 5.*)

Once the tracker performed satisfactory the user interface on the PC was developed and a communication link was established between the VFG board and the PC. One of the first features implemented was the recording of eye movements. The recording is initiated with a left mouse button press and stopped with a second press on which the centroid co-ordinates are downloaded to the PC. In one of the first tests a person's eye movements was recorded while tracing the outline of the PC monitor. Figure 3.5 shows the recorded eye movements.

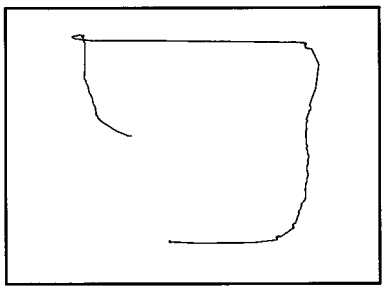

Figure 3.5 First Recorded Pupil Centroid Co-ordinates.

Next the purkinje centroid was added. A fixation grid pattern was displayed on the PC monitor and the pupil and purkinje centroid positions were recorded for each fixation. The initial fixation grid contained 20 points that were uniformly distributed on the display. The subject had to fixate on each test point and pressed the mouse button when ready on which the pupil and purkinje co-ordinates were recorded and sent to the PC.

Figure 3.6 shows the calibration pattern on the left and the recorded pupil *(crosses)* and purkinje *(circles)* positions on the right.

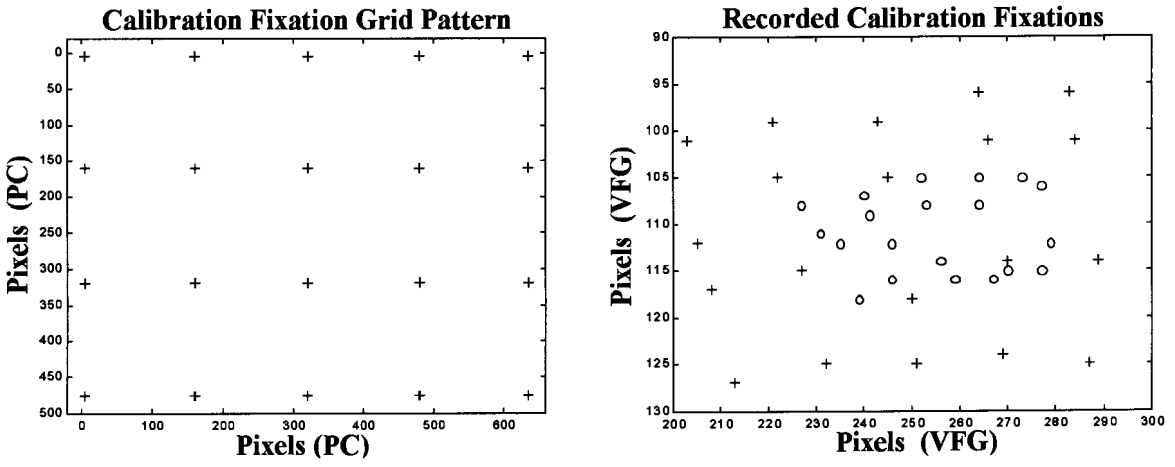

Figure 3.6 Fixation Pattern and Recorded Pupil and Purkinje Centroids

In order to determine the gaze position on the monitor the PPD features needed to be mapped to the displayed grid positions. Thus the difference between the leftmost cross and leftmost circle indicated on the right of figure 3.6 had to be mapped to the top left fixation co-ordinate on the display. Note the skewed and nonlinear relationship between displayed data and the fixation grid. The above process is the typical approach used in various systems where gaze calibration is accomplished by recording a sequence of user fixations, stimulated via fixation points on a computer display [19], [29], [30], [31],[32].

In Sheena's work in [23] a transform, containing 14 coefficients, is used to determine the relation between the eye features and the scene. The 14-term system is broken down into several smaller systems, containing only a maximum of three equations that need to be solved. Each small system has a particular function, e.g. supplying the estimate horizontal eye position, vertical eye position, head motion crosstalk correction and mirror corrections.

This particular situation lends itself to a neural network (NN) implementation, which is ideally suited for mapping and interpolating a given set of input data to its associated output data, without requiring specific knowledge of the system equations.

In Pomplun's work in [30] a "parameterized self-organising map" (PSOM) artificial NN is proposed for an eye tracker implementation. An adaptive NN is used to compensate for measurement errors of the system. This method can be trained quite fast with a very small number of training examples. This process is however not yet integrated into a real time system.

A feedforward NN configuration [34], trained via backpropagation, was implemented in the prototype platform. This architecture was chosen for its simplicity; it is well known and can easily be adapted for embedded applications. The Levenberg-Marquardt optimization technique [33] was used the in the learning procedure for faster training results.

> "Trained backpropagation networks tend to give reasonable answers when presented with inputs that they have never seen. This generalization property makes it possible to train a network on a representative set of input/target pairs and get good results without training the network on all possible input/output pairs." [33]

In order to develop and test the initial NN procedure, three different fixation test patterns were recorded for about 15 subjects. The first two sets were used for calibration purposes and the third set was used to evaluate the performance of the mapping process. The results obtained with the prototype platform verified the conceptual design of the system, but did not exhibit a robust and reproducible output.

It must be remembered though that the performance of the initial NN implementation is directly coupled to the feature extraction accuracy. *(The results obtained at this stage are discussed in chapter five where the algorithm implementation and detailed NN implementation are described.)* Although the results did not comply with the given specifications, we were satisfied that with some refinement and with the more optimal SFPD layout, a RMS accuracy of 1° could be achieved.

Chapter 4

SFPD Concepts & Experimental Setup

The results obtained with the prototype apparatus and beta algorithms verified the validity of the initial design decisions. With these results the layout and integration of the eye tracker design for the SFPD system could be confirmed. The experimental apparatus configuration used in the SFPD is described in the third section of this chapter.

However, before we continue, we first need to give a more detailed description of the SFPD concept, which is described in section two. In the first section some background is given on why a SFPD is needed. The information regarding the SFPD concept has been gathered from Viljoen's work in [2].

4.1 Tele-existence

In hostile environments like space, nuclear or other unstable sites, remote operation plays an important role. Sheridan [1] defines a teleoperator as a machine that operates on its environment and is controlled by a human at a distance. The human operator often uses displays and controls to command the teleoperator. In Tachi's [3] work a tele-existence system is designed and evaluated. In his work he states that one of the problems that must be solved for an ideal remote operation system is: "to present the human operator with sensory information of the slave robot's environment as naturally as possible. "

In the introduction it was stated that monoscopic vision is required in a far-field teleoperator and in chapter two it was indicated that foveal vision is needed for detail scene analysis. Peripheral vision however, is very important in search performance tasks. Studies have shown that search time and peripheral acuity are inversely related [6]. In studies that examine search efficiency or the relationship between search and visual acuity the concept of the visual lobe is used [17], [26].

The visual lobe is defined as the limit of peripheral acuity for a particular target and background; and is related to the probability of detecting the same target in a single eye fixation during search [17]. The size of the visual lobe is unique for each person and can be increased with training. However, for more difficult visual tasks, foveal acuity is a better predictor of detection capability than peripheral acuity [7]. Thus, to induce telepresence in a far-field teleoperator one needs both peripheral and foveal vision. If one only uses a narrow field display, detection capability would be severely limited and will result in tunnel-vision search.

The ideal head-mounted display configuration would therefore have very high-resolution, wide field displays that extend over the entire peripheral field. The displays would be slaved to the operator's head movements and would be updated continuously to ensure a total wrap-around scenario to induce the feeling of physical presence. Figure 4.1 depicts this scenario: a high-resolution display that covers the peripheral field-of-view (FOV); conveying peripheral vision, and providing detailed images at fixation points.

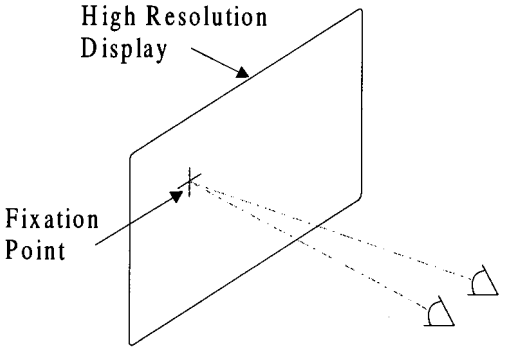

Figure 4.1 Wide field, high-resolution display conveying both peripheral and foveal vision.

As alluded to in the introduction, this type of display is not technologically feasible at present. To overcome this limitation some systems alternate between a wide and narrow FOV, which degenerate the natural aspect of the interface and hampers detection performance.

According to Yamaguchi precise pixel expression and a large visual field display are required for enhanced reality. Unfortunately this combination is contradictory, because the view angle per pixel deteriorates with an increase in the size of the visual field. In order to solve this

problem he proposes a system where a high-resolution central vision image is optically superimposed on a large peripheral image. This concept is depicted in figure 4.2.

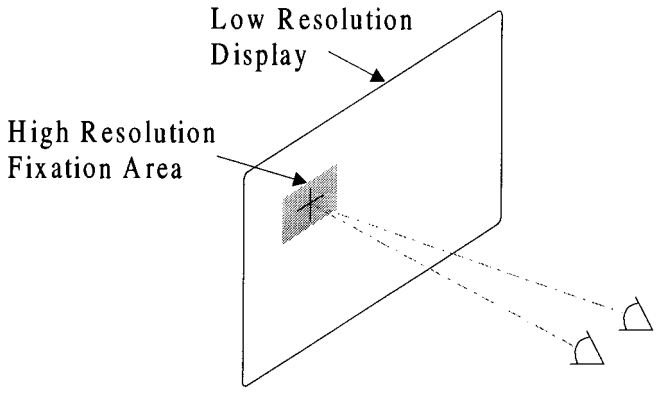

Figure 4.2 Wide field, low resolution display with a superimposed high resolution image at the gaze position.

This technique exploits the decrease in visual acuity of the eye towards the periphery in that the fovea is quite small; subtending a visual angle of about one degree [14]. In this concept an eye tracker is needed to determine the gaze position for the high-resolution display. Yamaguchi demonstrated the feasibility of this concept by employing a large screen and image projector for the wide field display and an optical image positioner for the foveal image. Figure 4.3 shows the system configuration, with a more detailed view of the projection implementation on the right.

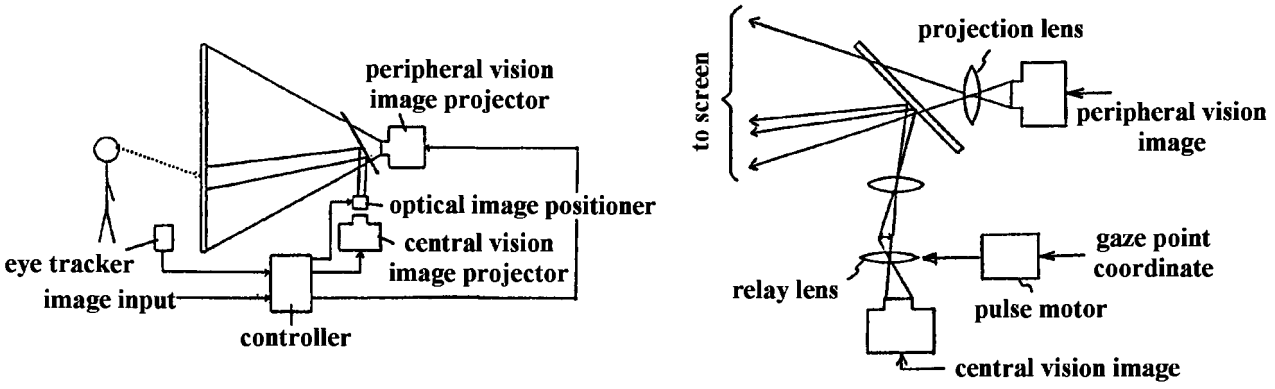

Figure 4.3 Configuration for the large visual display, with its projection implementation. [9]

Yamaguchi also proposed a reconfiguration of the projection system for a head mounted display type system, which is shown in figure 4.4.

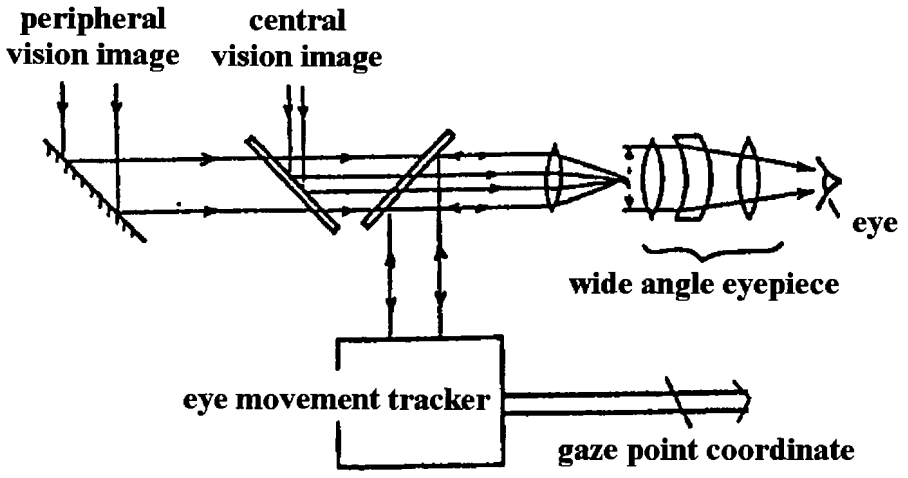

Figure 4.4 Yamaguchi's proposed projection configuration for a head-mounted display. [9]

The system proposed in figure 4.4 is quite complex and it would be difficult to combine the optical projection mechanism and eye movement tracker into a compact design. In this type of design a single projection display can be used, but additional optics would be required to physically reduce its size and to ensure adequate peripheral coverage. Using a display for each eye would unnecessarily complicate the design in that a foveal projection image would be required for each eye.

In his work, Viljoen proposes the SFPD concept that elaborates on the above idea. His approach provides an elegant solution that results in a much simpler implementation.

> "*A display which combines both a foveal and peripheral display channel require an extremely complex optical design. The display can be greatly simplified if the peripheral and foveal display channels can be split between the two eyes of the operator, with the human visual system fusing the picture.*" [2]

4.2 Split-Foveal-Peripheral-Display Concept

In the SFPD concept a narrow field, high-resolution picture is projected into one eye and a wide FOV, low-resolution picture into the other eye. The human brain fuses the two images into one coherent picture. In [2] Viljoen analyses the conditions required for the brain to fuse both images into one coherent picture and proves the feasibility of the concept. This process is by no means an artificial stimulation and is often find in people with a dominant eye, where the brain adapts and uses the dominant eye for detailed vision and the other eye mainly for peripheral vision.

The device used to test the concept is shown in figure 4.5. The device is fixed to a windowpane where a target can be viewed at infinity. The large cutout simulates a 60° peripheral FOV, with a small cutout in the center to block out foveal vision. On the right side only foveal vision is allowed via the small cutout. The user, positioned behind the tester fixates on the distant target with each eye aligned with its central cutout. Thus a narrow field image is received in his right eye and a wide field image in his left eye, which the brain then fuses into one monoscopic image. Ten subjects were tested and each one was able to fuse the two pictures, with two of the subjects needing some training before they could fuse the pictures successfully.

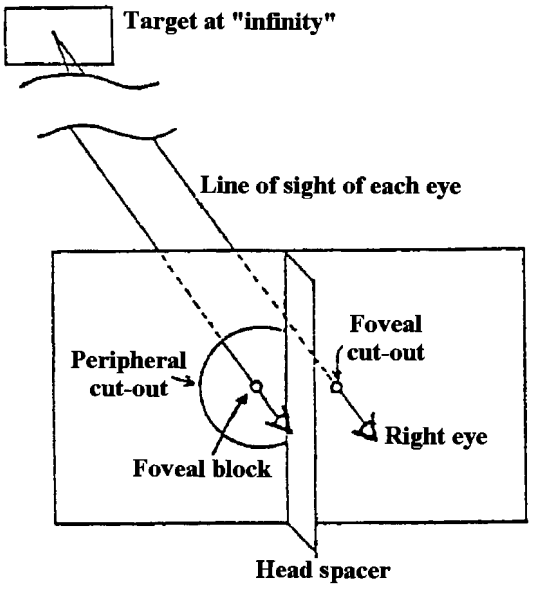

Figure 4.5 The SFPD concept tester. [2]

With the basic concept proven Viljoen built an electro-optic SFPD system to analyse the fusion phenomenon within typical teleoperator conditions. The experimental setup consisted of two CCD cameras, one with a wide field lens and the other with a narrow field zoom lens. The user viewed the images on two CRT displays via two eyepieces, mounted on a structure that allowed several adjustments.

Fusion occurred for all subjects if the alignment is within 1.5° in azimuth and 0.4° in elevation, with the magnification matched within +10% and –20%. It was shown that the SFPD gives both peripheral and foveal vision within a teleoperator scenario and that the concept is realizable within current technological constraints. Figure 4.6 depicts a typical teleoperator configuration with the NF and WF cameras slaved to the user's head movement.

The SFPD concept provides an elegant solution for presenting both peripheral and foveal vision in a head-mounted display. In the SFPD system we rely on the brain to impose/fuse the foveal image within the peripheral image.

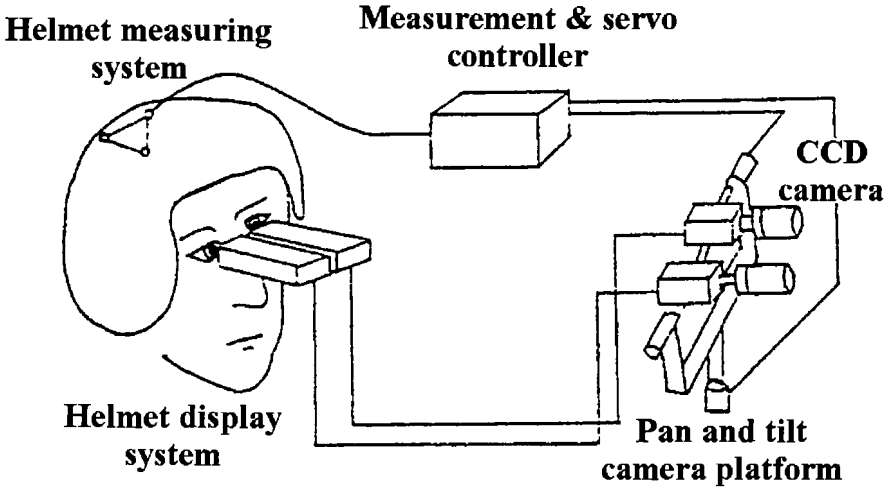

Figure 4.6 SFPD teleoperator configuration for tele-existence. [2]

4.3 SFPD Experimental Setup

In this section a brief overview is given of the SFPD experimental apparatus. A detailed description of the system configuration can be found in Viljoen's work in [10]. As pointed out earlier, the prototype platform was used to determine and verify the eye-movement recording procedure and configuration.

The wide FOV optics is fixed, but the NF or foveal image must be projected at the operator's gaze direction. This is accomplished by means of two servoed mirrors which projects the display in azimuth and elevation respectively. In the SFPD system the CCD camera is placed in front of the eye, as opposed to the prototype setup where a mirror is used to reflect the IR image into the camera. In this instance the dichroic mirror reflects the visible CRT display image and passes IR light from the illuminated eye to the camera. Figure 4.7 shows the layout of the NF and WF optics.

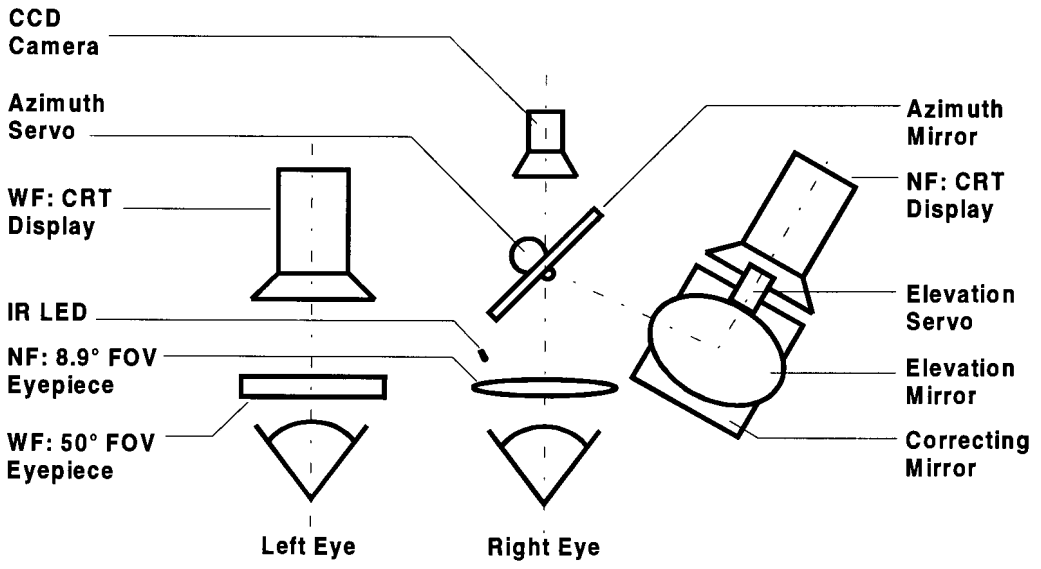

Figure 4.7 Optical layout of the eye-directed SFPD [10]

Initially the IR LED was positioned on the optical axis, in front of the camera, but at this location it did not provide enough illumination at the eye and was relocated as shown. The VFG board used in the prototype platform receives the captured eye images, determines the gaze direction and directs the mirror servo units via 12-bit serial commands.

These commands are also used to direct similar servo units on the external NF image camera. During the test phase the system was used in simulation mode and a pre-recorded veld scene was used and updated via a NF and WF simulator PC. Viljoen captured the images, using the external WF and NF cameras, and painstakingly converted the video images into a virtual scene; using the servo commands to correctly update the displays. A block diagram of the experimental setup is shown in figure 4.8. The highlighted blocks indicate the area targeted in this dissertation. The DSP interface software, calibration routines and test software developed in this dissertation are located on the NF simulator PC.

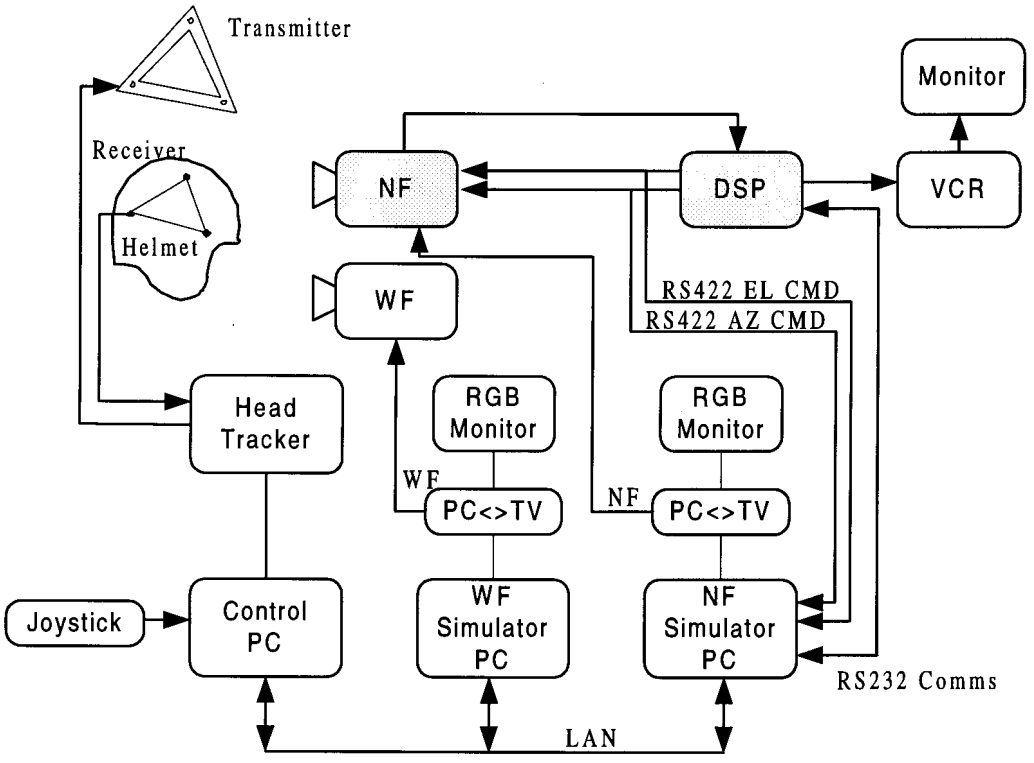

Figure 4.8 Eye-slaved helmet display system block diagram. [10]

Figure 4.9 on the next page depicts a summarized view of the components related to this study and the interaction between them. The frame grabber DSP unit extracts the PPD from the CCD camera video frames, resolves the gaze position and issues the mapped azimuth and elevation commands to the servo units and to the NF computer. The NF computer updates the CRT display with the new high-resolution image at the given gaze position. In parallel the mirrors are positioned via the servos to project the foveal image at the gaze position. All of

this occurs in less than 100 milliseconds, in time to have stabilized when vision exits its inhibitory phase.

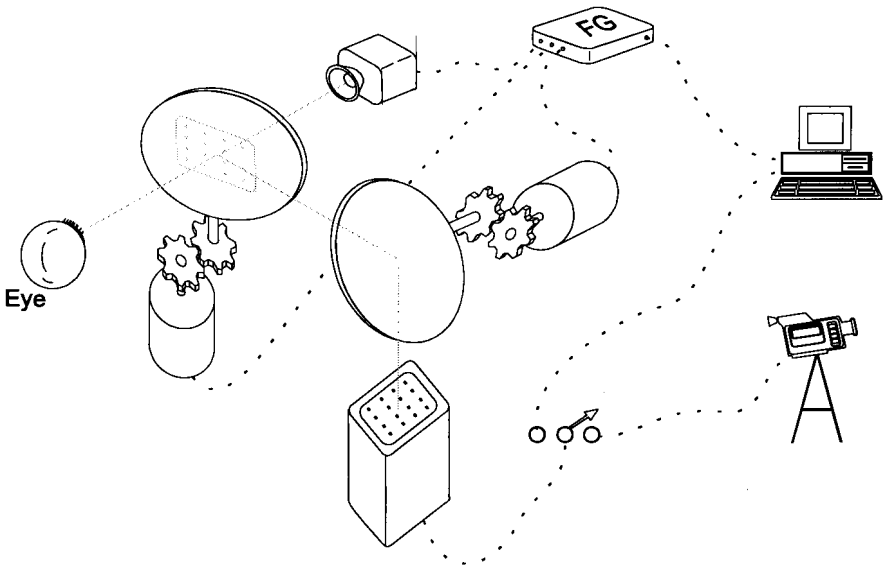

Figure 4.9 Eye-directed system configuration.

A photograph of the full system experimental setup is shown on figure 4.10. The monitor on the far right shows the eye tracker's feedback image. Next to the feedback monitor is the WF computer with the WF image on its display. The DSP unit is behind the WF computer, obstructed from view. Next is the NF computer with its display and on the left the control computer. Behind the operator and the head-up display unit one can also catch a glimpse of the head tracker mounted against the wall.

The choice for the specific layout of the NF channel design will be described in the next chapter. Although the apparatus is quite bulky and awkward in its initial phase, it enabled us to perform various tests and verify the conceptual design and tentative assumptions.

Figure 4.10 SFPD system experimental setup. [10]

Chapter 5

Conceptual Design & Methodology

The previous chapters covered the groundwork and nomenclature needed in the development of the eye-directed controller. A prototype system was developed that demonstrated the basic requirements for an eye-directed controller and proved the rationale behind the neural network implementation. Although the feature extraction routines and neural network implementation showed plausible results, some refinement was needed to ensure reproducibility and robustness. In the previous chapter the SFPD experimental configuration was presented, which described the optical design and mechanical operation of the system.

In this chapter the conceptual design followed in providing an eye-directed controller for the SFPD system is given. The methodology used in the feature extraction and neural network routines is described. Some novel image feature extraction techniques that have been developed for more robust, real-time feature extraction are discussed in the second section. The proposed neural network implementation architecture is discussed in the third section.

With reference to the previous chapter the SFPD system can be depicted as follows:

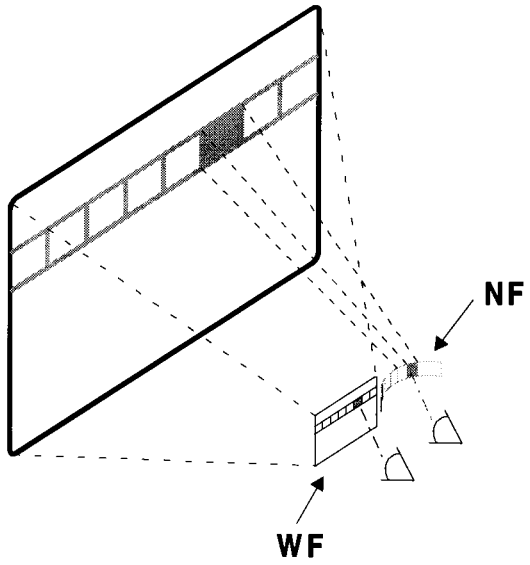

Figure 5.1 SFPD system with WF, low resolution display at the left eye and the high resolution NF display at the right eye.

Figure 5.1 depicts the peripheral scene with the imposed foveal image at the gaze position. The WF display projects the peripheral image into the operator's left eye and is updated with head movement. The NF display at the right eye must be projected at the gaze position by means of the servoed mirrors, and needs to be updated with any change in the gaze position.

The controller must determine the NF projection position, by calculating the gaze direction of the operator, using the extracted pupil-purkinje-difference (PPD) components. This process is depicted in figure 5.2. The NF position at the bottom-right is determined from the gaze position, indicated by the highlighted rectangle in the WF domain, which in turn is determined from the PPD shown on the left.

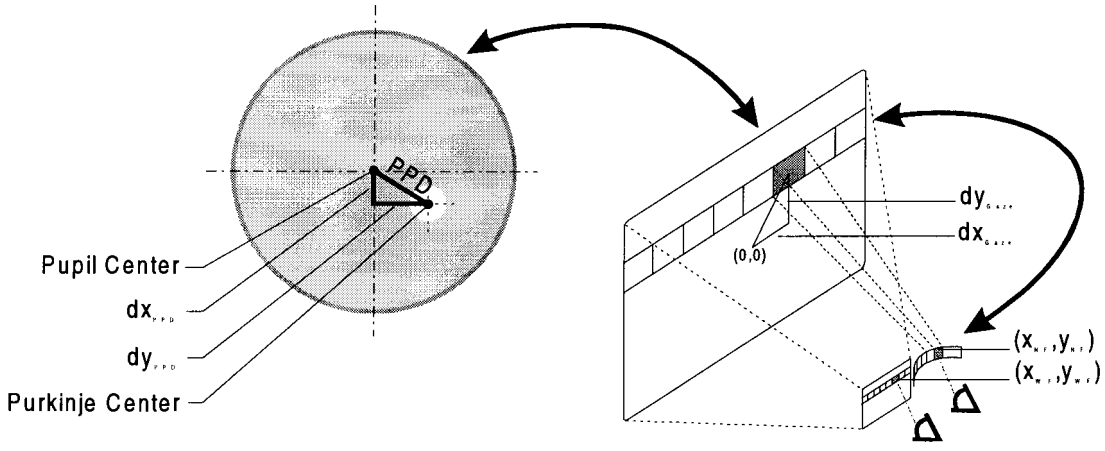

Figure 5.2 Graphical representation of the eye-directed controller functionality.

5.1 Conceptual Design

Given the extracted PPD features, these features need to be mapped/transformed to the gaze position. In turn the gaze position needs to be mapped to the servo positions for the correct NF projection.

$$\textbf{PPD}(dx,dy) \quad \Leftrightarrow \quad \textbf{Gaze}(dx,dy) \quad \Leftrightarrow \quad \textbf{Servo}(dx,dy)$$

There is a fixed gain setting in the WF and NF image alignment for a particular operator, or a fixed gain between the WF and NF cameras that captures the distant images. However, the projected NF display position is unique for each subject, thus each subject will have a different servo setting for a particular fixation point, which in turn necessitates an additional mapping between the point of gaze and the servo positions.

As the NF channel is already cluttered with the projection mirrors, one intuitively wants to allocate the gaze determination process to the fixed WF channel. In this approach the gaze position could easily be determined within the WF display, quite similar to the setup used in the prototype apparatus. An additional transformation process is now required to map the WF gaze position to the respective servo command positions. The transformation processes required in this approach are:

$$T_1(dx_{PPD}, dy_{PPD}) \quad \Rightarrow \quad (x_{WF}, y_{WF})$$

$$T_2(x_{WF}, y_{WF}) \quad \Rightarrow \quad (x_S, y_S) \quad \Leftrightarrow \quad (x_{NF}, y_{NF})$$

As indicated already, the first transformation process can be accomplished via the neural network (NN) implementation used in the prototype design. In this approach a fixation grid shown in the WF display would be used for calibration/training purposes. Similarly a NN could also be used for the second transformation, but in practice this approach would require a very tedious calibration process.

Fixation grid points would be used on the WF display to record the relevant PPD features needed for solving T_1. In parallel with this the servo positions would have to be adjusted to ensure fusion of the images at each fixation point in order to solve T_2. This can only be executed with feedback from the operator and would drastically slow down the calibration process. Prior to these procedures the WF-NF gain ratio for the cameras must be determined to ensure proper alignment for fusion. Another factor that can induce errors in this approach is disjunctive eye movements or phoria [40]. "Phoria represents the degree to which the ocular axes verge in front of, at, or beyond a given target in the absence of fusional stimuli."[40]

In this dissertation we propose a different approach, which elegantly solves the required transformation processes.

If one could directly map the PPD parameter to the servo commands, only a single transformation process would be needed. This can be accomplished if the eye tracker is mounted in the NF channel. By displaying a fixation point at different gaze positions or servo positions within the NF, instead of using the static displayed fixation grid, one directly compares the gaze position with the PPD parameters. With this approach the second transformation process is eliminated. Thus, we have:

$$T_1(dx_{PPD}, dy_{PPD}) \quad \Rightarrow \quad (x_S, y_S) \quad \Leftrightarrow \quad (x_{NF}, y_{NF})$$

During this process the WF display can be blanked. Once the calibration has been completed the WF image can be activated and the WF-NF gain adjusted by simply showing the NF at two successive locations, and adjusting the gain for fusion with feedback from the operator. Thus, rather than determining the gaze point in the peripheral scene and then resolving the foveal positioning; one can determine the foveal gaze direction in space and impose this on the peripheral scene.

Via this approach the raw 12-bit control commands are directly mapped to the gaze position co-ordinates and not a single axis transformation is required.

5.1.1 Functional Layout

Functionally we can depict the eye directed controller as shown in figure 5.3. The two main functional units that are addressed in this chapter are:

- Image feature extraction algorithms for PPD feature extraction.
- Neural network gaze mapping.

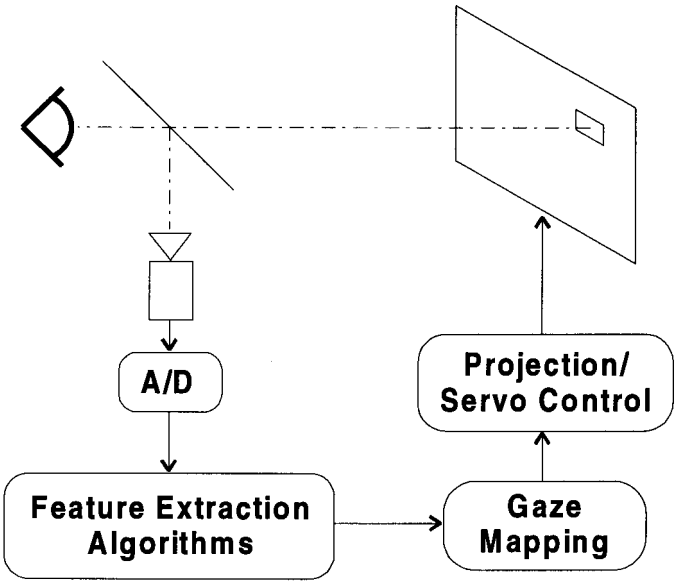

Figure 5.3 Functional layout of the eye-directed controller.

An interesting analogy between the stages found in the visual pathway and the stages in the eye-directed controller is shown in figure 5.4.

As mentioned earlier, one of the main constraints in the design process is the real-time processing requirement. The specified total delay time in the NF display update is 100 ms.[10] This delay time consists of the eye-directed controller response time, the servo response time and the image update response time. The eye-directed controller has a delay time of 20-40 ms, which is determined by the video frame rate (*20 ms*).

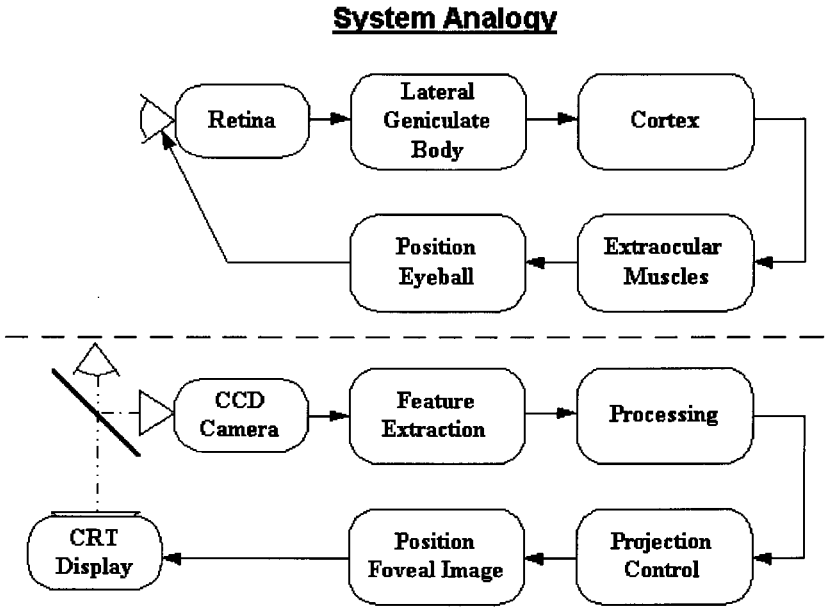

Figure 5.4 Analogy between the visual pathway and eye-directed controller stages.

As indicated in chapter 2, the minimum refractory period between saccadic eye movements varies between 100-200 ms. Therefore the NF display needs to be stable and correctly positioned within 100 ms.

5.2 Image Feature Extraction Algorithm Development

One of the main functions of the DSP block is to extract and track the pupil and purkinje coordinates from each image frame. The pupil and first purkinje image coordinates are defined as the centroid positions, or center positions of the pupil and purkinje areas respectively. As described in chapter 3, the difference between the pupil and purkinje coordinates serves as an indication of the gaze direction. Once these coordinates have been extracted, it can be mapped to the person's gaze position within the visual domain. The features that must be extracted via the image processing algorithm serve as input to the NN and will therefore have a direct effect on the overall system performance.

In essence, an image processing algorithm must be developed that can track eye movements within the following criteria:

- Real-time capability
  Comply with the computation time constraint.
- Accuracy and robustness
  Ability to handle distortions, including occlusions and changes in light conditions.

There are various image processing techniques available that can enhance the pupil image and pinpoint its centroid with sub-pixel accuracy, even with substantial occlusion. Unfortunately the real-time constraint drastically limits the choices. As stated in the previous section, the total control loop has a 20 ms timeframe, which includes the mapping process. To give some indication of the severity of this constraint: by simply comparing each pixel in a frame to a set threshold level, the allotted time for the image processing phase would be exceeded. An innovative technique must therefore be developed that provides simple and stable feature extraction characteristics.

In Wagner's thesis [15] image registration based techniques, using template matching, are evaluated for eye movement measurement. Instead of using computational costly correlation techniques Wagner opted for more economical similarity measures. The results obtained look promising, but a full template matching technique would simply exhaust the available time in

the current configuration. An improvement on computational cost on Wagner's techniques would be to implement a fast search technique like the modified *n*-step search technique presented in Dufaux's [35] paper. This technique can reduce the number of match positions by two orders of magnitude and both the mean absolute error and mean square error measures can be used.

In Rubenstein's work [36] several image centroid algorithms are evaluated. In the initial development it was decided to use this centroid algorithm to determine the pupil and purkinje centroids. This technique is described in the first section, along with the pitfalls encountered during implementation. In order to overcome the occlusion artifacts that plagued the centroid technique, two novel techniques have been developed in the search for an enhanced and more cost-effective algorithm. These techniques are described in the second section. Although these techniques exhibited their own inadequacies, it paved the way for the final bisection algorithm that is presented in the third section.

During testing, the extracted features are indicated on the video output. This process serves as an immensely valuable debugging aid. Through this feedback information the stability and accuracy of the algorithm could be checked immediately. Optimal settings of various parameters were achieved via this feedback process.

In retrospect, it can be said that most of the time spent to obtain a robust and accurate eye-directed system, can be linked back to the efforts in searching for a solid feature extracting technique.

5.2.1 Centroid Tracker Implementation

In Pratt [37] the centroid of an image is defined as the ratio between the first-order and zero-order spatial moments. If $F(j,k)$ represents the image, the centroid position (x_k,y_j) is:

$$\bar{x}_k = \frac{M(1,0)}{M(0,0)} \qquad \bar{y}_j = \frac{M(0,1)}{M(0,0)}$$

$$where \quad M(0,0) = \sum_{j=1}^{J}\sum_{k=1}^{K} F(j,k)$$

$$M(1,0) = \frac{1}{K}\sum_{j=1}^{J}\sum_{k=1}^{K} x_k F(j,k)$$

$$M(0,1) = \frac{1}{J}\sum_{j=1}^{J}\sum_{k=1}^{K} y_j F(j,k)$$

The centroid (x_k,y_j) gives the balance point for $F(j,k)$. In order to determine the centroid positions for the pupil and purkinje areas in the image, both areas must be isolated or segmented. At a first glance, the pupil and purkinje areas have distinct intensity values, which can be used for segmentation purposes. As the pupil area absorbs the infrared light, it appears black and has the lowest intensity values in the image. The purkinje area in turn appears bright white and has high intensity values. From the sampled infrared image shown in figure 3.4 the high intensity purkinje reflection is clearly visible as a white circular spot. At a first glance it looks if the areas can easily be segmented via fixed intensity threshold settings.

In using a segmenting technique the image is reduced to a binary image, in which the zero-order moment is equal to the image area. This effectively describes the center of gravity algorithm in Rubenstein's [36] thesis where the subtracted threshold level is set at the continuous background level. This technique was evaluated as the most accurate centroid algorithm in his tests and was implemented in the final centroid algorithm via the following equations:

$$M = \sum_{j=1}^{J}\sum_{k=1}^{K} T(j,k)$$

$$x_k = \frac{1}{M}\sum_{j=1}^{J}\sum_{k=1}^{K} x_k \cdot T(j,k)$$

$$y_k = \frac{1}{M}\sum_{j=1}^{J}\sum_{k=1}^{K} y_k \cdot T(j,k)$$

$$\text{where} \quad T(j,k) = \begin{cases} 1 & F(j,k) > Threshold \\ 0 & Otherwise \end{cases}$$

In order to do real-time tracking of the centroid positions a search or tracking window/mask is used to reduce calculations. Initially the search window size is set to the image size and the image is scanned for pixels lower than the fixed pupil threshold. To reduce the computation time only every second row, and every fifth pixel in the row is checked with the fully expanded mask. Once the centroid positions are determined the search mask is shrinked to a border window encompassing the eye, which drastically reduces the number of pixels to be searched. To avoid detecting spurious small regions that satisfy this criterion, like the corner of the eye, a fixed number of pixels must detected within the search region. A circular border window was used, as shown in figure 5.5. With an eye blink, or with loss of tracking, the search window is expanded again. Note that one of the biggest advantages of the centroid technique is its sub-pixel accuracy.

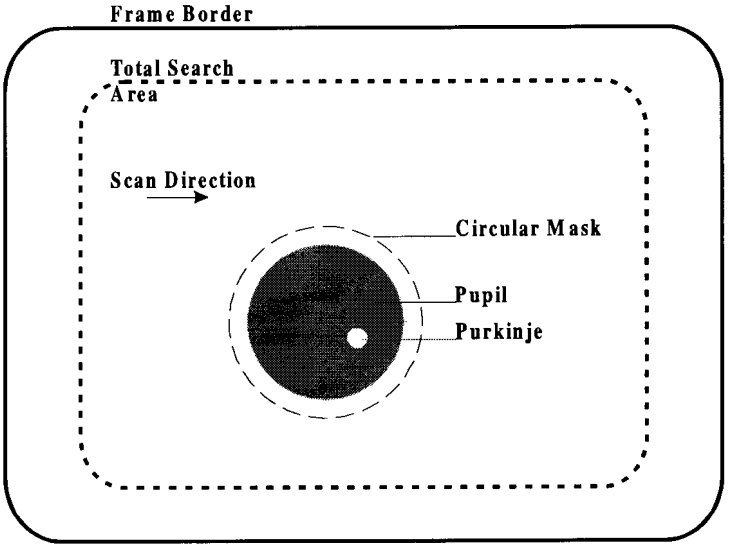

Figure 5.5 Tracking window in centroid technique implementation.

The tracking algorithm was initially implemented with fixed threshold levels that resulted in occasional track-loss and very jittery output positions. The tracking performance also varied drastically between different subjects. At closer examination the following problems were identified:

- With the IR LED focused on the eye corner, the image intensity values fade to the right of the image, resulting in lower intensity values on the right pupil border. This makes it much more difficult to choose a fixed threshold value for pupil and iris separation.

- The average intensity values differ between persons and individual settings were required for optimal performance. The intensity varies with the amount the eye surface reflects light. Here the setting of the eye-sockets, especially deeply set eye-sockets caused huge variations.
- Looking down, jittery outputs occur when eyelashes enter the pupil area. Eyelashes usually appear brighter and are therefore excluded in the centroid calculations via the pupil threshold. The upper part of the pupil is occluded with maximum downward gaze. In such cases the centroid is simply the center of the visible pupil region.
- The CCD camera has automatic exposure and gain control. With different gaze positions the amount of reflected light changes and sometimes triggers a new control setting of the camera, which in turn changes the intensity values.
- With subjects wearing glasses and in some cases, hard contact lenses, secondary bright LED reflection images appear and make purkinje image tracking impossible.
- Very small, vertical jitter occurs because of interlacing between even and odd video frames. With some filtering in later stages of the system, this jittering is negligible.

The intensity variance is a result of the IR LED position. As mentioned in chapter 4, the IR LED is situated on the side of the NF apparatus. Figure 5.6 shows the typical radiation characteristics of a GaAs IR LED. The focus position of the LED is pointed at the eye corner to inhibit glare reflections from the anterior part of the eye. The eye is therefore not evenly illuminated and the light intensity disperses towards the other eye corner. This effect is aggravated when the gaze position is directed away from LED focus point, resulting in a darker eye image, and vice versa to the other side.

The purkinje spot consistently has very high intensity values and its centroid can be extracted without any difficulty. Although the iris and pupil areas on the image are easily distinguishable, the intensity differences between them change with position on the image and are much smaller in comparison with the pupil-purkinje intensity difference. As the points above indicate, the biggest problem is to distinguish between the iris and pupil intensities, which necessitate dynamic or adaptive thresholding techniques.

With the current experimental setup the purkinje area never leaves the pupil area, except in extreme cases with certain subjects when their gaze direction is to the far right. Even then the outside borders of the two areas touch.

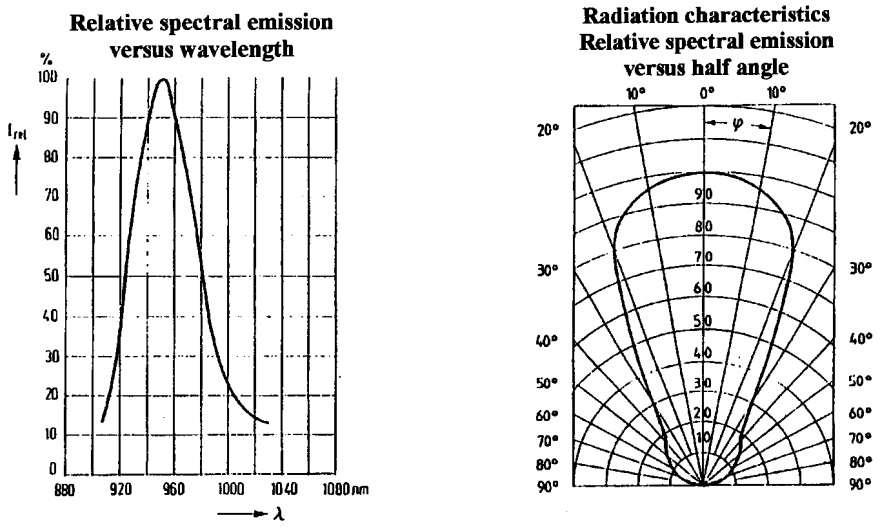

Figure 5.6 Radiation characteristics of a typical IR LED.

This effect lead to the following dynamic threshold setting technique:

1. First determine the purkinje position in the image. (Search for intensities > $Th_{purkinje}$)
2. Sample the intensity value 20 pixels to the left of the purkinje position and set Th_{pupil} to this value. (*Given the constraints above, the sample point will always be in the pupil area.*)
3. Position the area mask at the sampled point.
4. Calculate the centroid position using Th_{pupil}. (Use a hysteresis of two intensity levels on Th_{pupil}.)
5. The purkinje area must be added to the pupil's area for accurate pupil centroid calculations.
6. Position a crosshair at the pupil centroid position for evaluation purposes.

These changes solved the loss of track problem and drastically improved the results with different subjects; no longer requiring individual intensity settings. Although the output jitter improved, jitter in the region of three pixels still remained, rendering the sub-pixel accuracy

A lowpass filter can easily solve the jitter problem, but the sub-pixel accuracy advantage is sacrificed. However, the biggest problem encountered in the centroid technique, is when eyelashes enter the pupil area, or the eyelid occludes the upper eye region and causes large offsets in the centroid positions. This only aggravates the non-linearity relationship of the PPD features.

The purkinje image is not obstructed and its position can constantly be extracted without any deviations. The pupil centroid however only represents the center position of the visible part of the pupil and deviates with the presence of any obstruction. From these results it is clear that a technique must be developed that is invariant to occlusions, like the template techniques presented in Walter's work [15]. The refined centroid technique provides a simple and accurate feature extraction technique, provided that there are no obstructions present. It provided the stepping stone in the development of the eye-directed system to implement and evaluate the mapping function and enabled validation of the system concept. It outlined some limitations of the system and surfaced unknown problem areas.

5.2.2 Enhanced Feature Extraction Algorithms

From the above it is clear that a technique had to be developed that was invariant to occlusions. In order to determine the centroid position of a partly occluded object template matching techniques can be used in which a pre-defined template is correlated with the image. The edges of the object can also be determined and interpolated in the occluded areas with pre-knowledge of the target shape. Unfortunately the real-time constraint prevents the use of these computation intensive techniques. Instead of using one of the formal techniques available, a new feature extraction technique was developed to address the specific problems within the system. In this section a brief overview is given on the first attempts.

5.2.2.1 Line-section-ratio

Figure 5.8 shows the PPD that must be extracted. The PPD can also be viewed as a ratio of its ***DX*** and ***DY*** components. With this idea in mind, together with the fact that we always have an unobstructed view of the purkinje image, a new technique was developed. The PPD

stays constant with small head movements, or stated differently: the first purkinje image position is fixed within the pupil area. Thus, if one could uniquely define the purkinje position within the pupil area, it would provide an analog to the PPD measure. Figure 5.9 shows the "Line-section-ratio" technique in which the DX and DY ratio uniquely defines the purkinje image position.

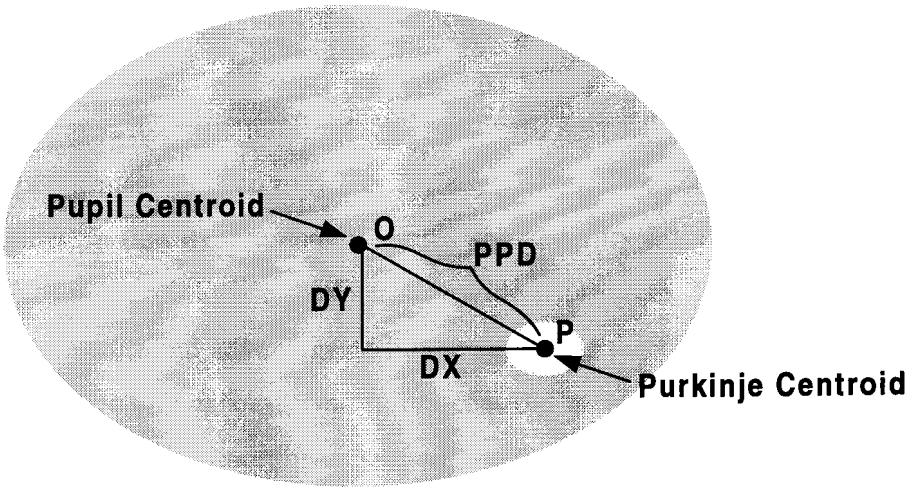

Figure 5.8 DX and DY components of the PPD.

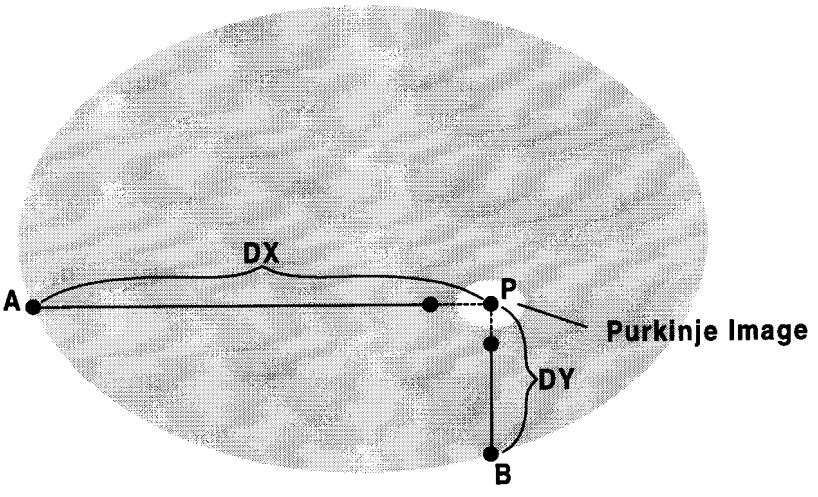

Figure 5.9 Line-section-ratio technique

We know that a circular template would be needed for template matching on the pupil. Instead of detecting the pupil's edges, one could detect three points on the perimeter *(instead of the previous two)* and determine the centre of the circle via the circle equation.

By using the "Equation of circle of radius R passing through origin" in [38] the following equations were determined: *(See Appendix A for the proof.)*

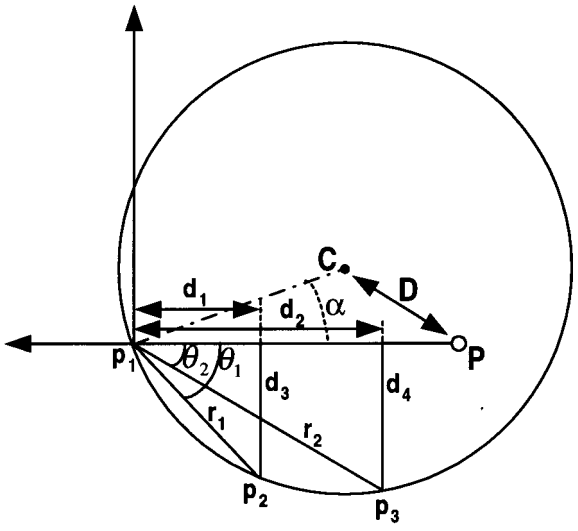

Figure 5.10 Obtaining centroid C via the perimeter point algorithm.

$$\overline{D} = \overline{C} - \overline{P} \quad where \;\; C \Rightarrow (R, \alpha)$$

$$\therefore \; D = \sqrt{(P_x - (p_{1_x} + R\cos\alpha))^2 + (R\sin\alpha)^2}$$

The above equation provides the PPD directly. Although the algorithm is more complex than the line-section-ratio technique, the pupil diameter problem is solved, as the radius of the pupil is taken into account in calculating the PPD.

The following steps were used in the perimeter point technique:

1. Determine the purkinje position ***P*** in the image. (Search for intensities > $Th_{purkinje}$)
2. Detect perimeter points p_1 and p_3. *(As in line ratio technique.)*
3. Follow the same p_3 procedure to detect p_2 at the projected midpoint between p_1 and p_3.
4. Determine the parameters required for equations 4.2 as shown in appendix C.
5. Finally calculate ***D*** to determine ***DX*** and ***DY*** of the PPD.
6. Position a crosshair at the pupil center position for evaluation purposes.

The results of the process are shown in figure 5.11. On the left the centroid is shown during the simulation phase, while on the right the result is shown on a captured frame. Notice that because of the interlacing between video frames, which results in a 512x256 digital resolution image, the captured eye shape in the image is elliptical.

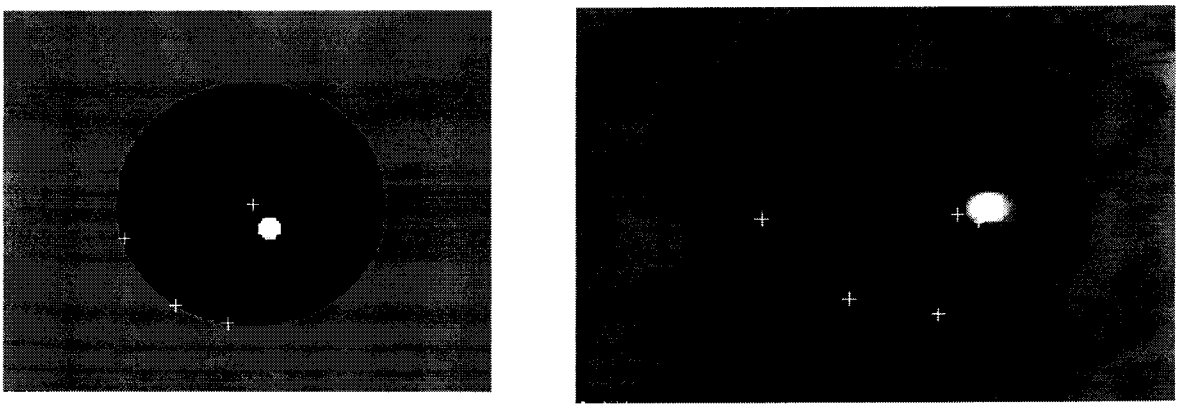

Figure 5.11 Perimeter point algorithm results.

Some corrections would have to be made to the algorithm to counter the elliptical representation, but even in the initial implementation stages the output showed large jitter artefacts and did not validate any further effort.

It was at this stage that the final algorithm, as described in the next section, came to light.

5.2.3 Bisection Feature Extraction Algorithm

In this section the final implemented feature extraction algorithm is described. This novel technique showed the best results within our configuration. Although theoretically it may lack the superb sub-pixel accuracy provided by the centroid technique; in practice it exhibits superior performance in both accuracy and robustness.

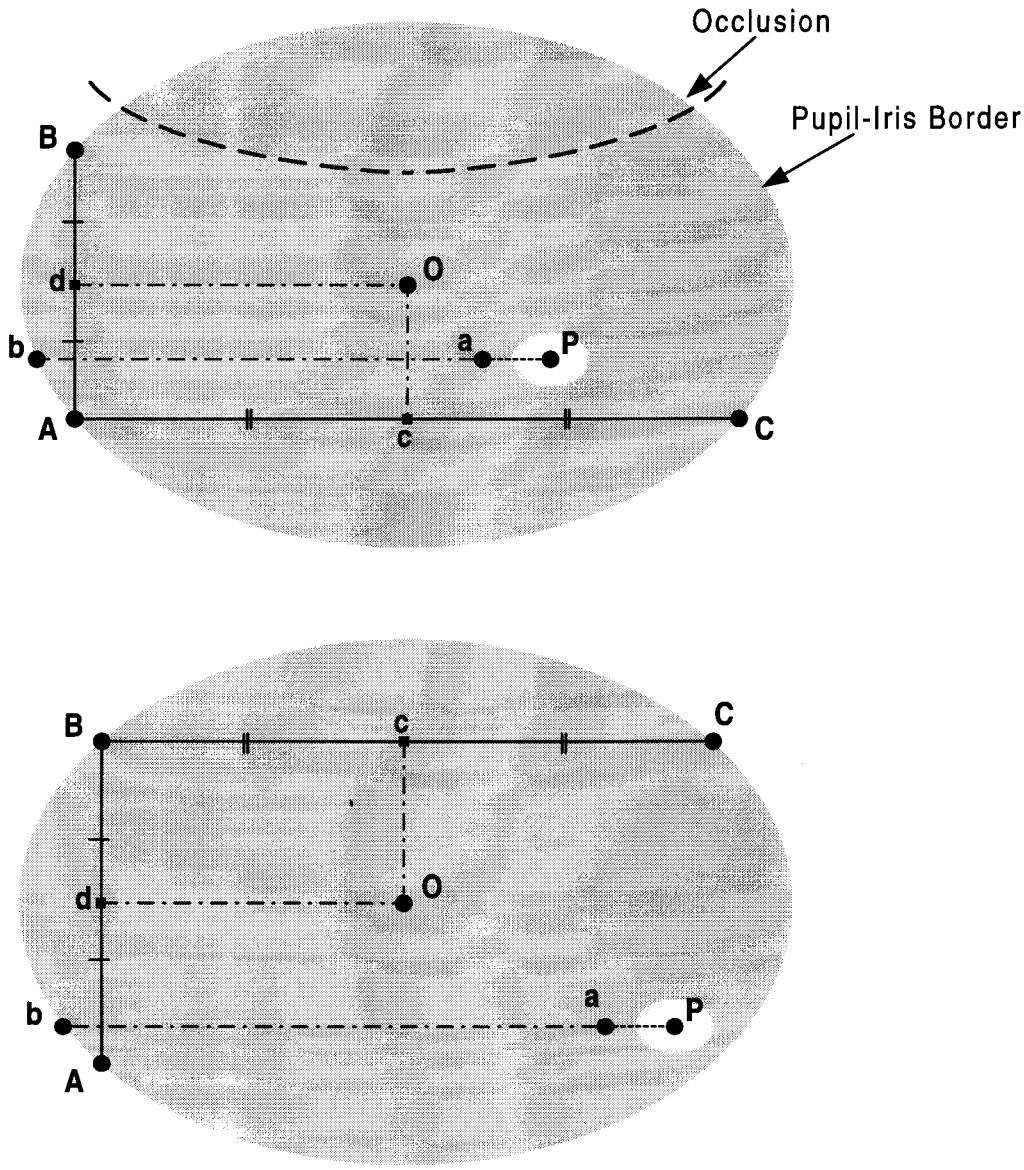

Figure 5.12 Bisection algorithm implementation.

The bisection technique provides a method for calculating the pupil centroid with the following characteristics:

- Invariant to the pupil diameter.
- Invariant to eyelash protrusions and upper region occlusions.
- Invariant to lighting changes and intensity variations.
- Invariant to elliptical deformation.

The steps in the bisection method are: *(See figure 5.12)*

1. Determine the purkinje position ***P*** in the image. (Search for intensities > $Th_{purkinje}$)
2. Detect edge ***b*** by scanning horizontally for an edge, starting 20 pixels to the left of ***P*** at ***a***.
3. Move 20 pixels inward from ***b***, scan upward for edge ***B*** and downward for edge ***A***.
4. If the vertical difference ***PA*** is greater than 30 pixels scan horizontally for edge ***C*** from point ***A***, else proceed from point ***B***.
5. The centroid coordinate ***O*** is then determined by bisecting ***AB*** for the vertical position and either ***AC*** or ***BC*** for the horizontal position.
6. The PPD inputs ***DX*** and ***DY*** are obtained from ***OP***.
7. Position a crosshair at ***A,B,C*** and ***O*** for evaluation purposes.

The bisection technique is a simple and effective method. The robustness of the technique is determined by the edge detection capabilities. When one looks upward the purkinje image moves toward the bottom of the pupil and the surrounding halo often interferes with the pupil border. The conditional step at step 4 ensures robustness in such instances, as depicted in the lower part of figure 5.12.

An optimal spatial mask filter has been incorporated for each scan direction. At first a simple difference mask was used, but at its optimal threshold setting it sometimes detected edges within the pupil. In one test subject a small off-center artifact in the pupil caused a faint bright spot in the image which was time and again detected, instead of the pupil border itself. The even and odd field alternations together with small lighting changes also caused small drift offsets in the edge detections.

The mask and assigned weights used for upward scanning is depicted in figure 5.13. Six pixels are used in the difference filter. The outside points inhibit detection within the pupil and ensure minimum drift at the pupil border. The cross indicates the current pixel in the scanning process. At the top pupil border the mask will indicate a large positive value and the threshold level can be set accordingly. Similar masks are used in the other scan directions.

In some instances small bright reflections occur in the eyelid region and can cause incorrect purkinje image detection. To avoid this problem the purkinje search starts at the right bottom corner. This also aids in expediting the purkinje detection process because most of the time the purkinje image is positioned in the lower half of the image. Finally the result of the bisection algorithm, as seen on the feedback monitor is shown in figure 5.14.

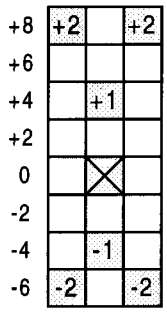

Figure 5.13 Spatial difference mask.

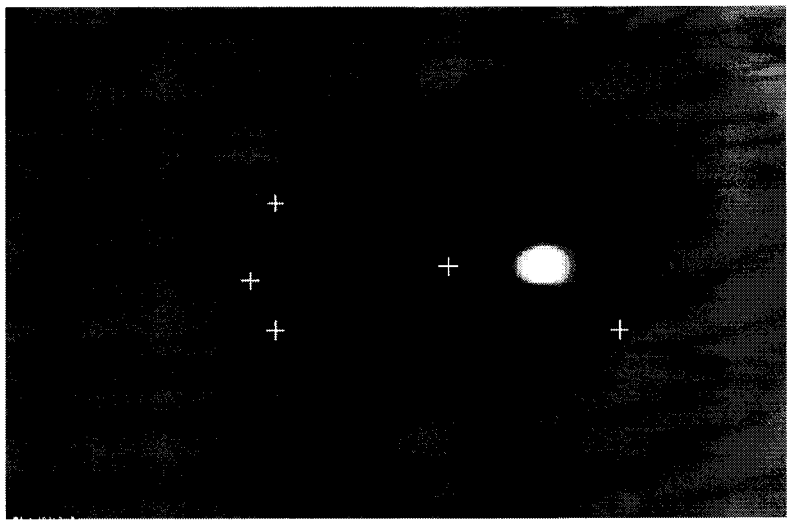

Figure 5.14 Output of bisection feature extraction algorithm.

5.3 Mapping the Gaze Position via a Neural Network

As mentioned in the conceptual layout and shown in figure 5.2, the gaze mapping functional block needs to map the extracted PPD features to the gaze position, which in turn is determined via the servo control commands. In chapter 3 the reasoning for the use of a feedforward neural network (NN) in the mapping process between the PPD features and the gaze position has been given and implemented. As alluded to earlier in this chapter it was suggested to implement a similar NN configuration in the SFPD system, which will directly map the extracted PPD features to the servo projection commands.

In this section the topology, calibration process and results obtained in the prototype NN implementation are described. Secondly the improvements and topology used for the SFPD NN implementation are discussed, and a description is given on the post-filtering techniques used on the NN output.

Note that a detailed description of the calibration process and NN training procedure used in the final setup, is given in chapter 6, as it is an integral part of the evaluation process.

The main aim in developing the NN is:

- The NN performance must be robust and repeatable for different subjects.
- The mapped gaze position must exhibit a RMS accuracy of 1°.
- Training must be fast and only the weights for each subject must be saved.

As suggested in chapter 3 the SFPD scenario lend itself to a NN implementation, which is ideally suited for mapping and interpolating a given set of input data to its associated output data, without requiring specific knowledge of the system equitions.

5.3.1 Prototype Platform Neural Network Implementation

In the prototype platform, described in chapter 3, the extracted eye features must be mapped to the gaze position on the PC monitor. Training of the NN on the captured data is executed on a PC for each user. The trained weights are saved and downloaded to the DSP board where the feedforward NN is implemented.

5.3.1.1 Topology

At first, instead of using the difference between the extracted centroid positions as input in the NN, the centroid coordinate positions were used directly. Figure 5.15 shows the structural layout of the prototype platform NN implementation.

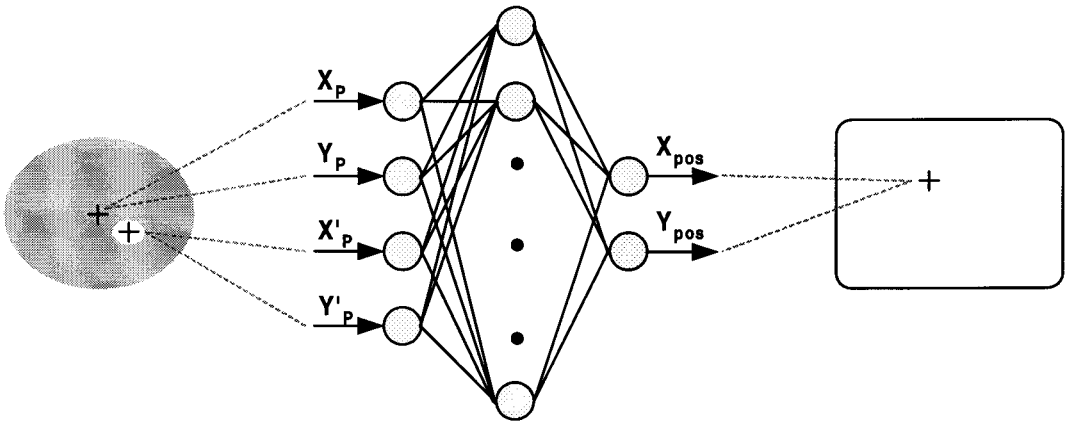

Figure 5.15 NN structure for the prototype platform.

A three layered structure is used, with four input neurons, a hidden layer and two output layers. The number of hidden nodes is determined by evaluating the NN output for fixation points excluded in the training set, checking for underfitting and overfitting conditions and whether the error threshold constraint is achieved within a fixed number of iterations.

The network is trained by epoch with randomly generated initial weights. An offset is added to the input data to ensure positive and negative inputs to the network that expedites the training phase. Tan-sigmoid transfer functions [33] are used in both the hidden and output layers to enable non-linear mapping. The input data is therefore scaled to include values

between –1 and 1. Tan-sigmoid transfer functions, instead of linear transform functions in the output layer resulted in better results and require offset and scale adjustments for the target data.

5.3.1.2 Calibration

In the calibration process a set of fixation points is sequentially presented to the user. The user must fixate on each calibration point and press the mouse button when ready. With each fixation the VFG card records the pupil and purkinje centroid positions. Once finished, the recorded centroid positions are sent to the PC and used in the training process.

Each set contains a fixed number of calibration points. Initially the points were presented in random order, but through various iterations it was found that the best results are obtained by presenting the calibration positions sequentially and restricting the movement per set to either horizontal or vertical lines. In this way the cross correlation between horizontal and vertical eye movement is suppressed. It seems that the apriori information of knowing the fixation presentation pattern, resulted in more stable fixation data. However, once the users are familiar with the calibration process, they tend to anticipate the next calibration position and shift their gaze position on pressing the mouse button. To inhibit this a delay time has been inserted between points. The test point flash period was altered to exhibit a 40-60 duty cycle, biased to the off period. With these enhancements the user must concentrate harder to fixate, while the flicker action stimulates visual attention.

Three sets were presented to the subjects, each containing 20 fixation points. Before the recording the user settles into a comfortable position, with the forehead pressed against the head restraint and the chin on the chin rest. Via the feedback monitor it is possible to ensure that the subject's eye is well positioned within the camera's field of view.

5.3.1.3 Results

The calibration pattern for each set is shown on the top right of figure 5.16. The crosses depict the first set which is sequentially presented in horizontal lines, the asterisks the second set, scanned vertically and the circles the third set which is used as test set. As can be seen in the second set, some calibration points are repeated in the set to increase the generalization of the network. Calibration sets for more than fifteen subjects have been recorded to develop and test the NN.

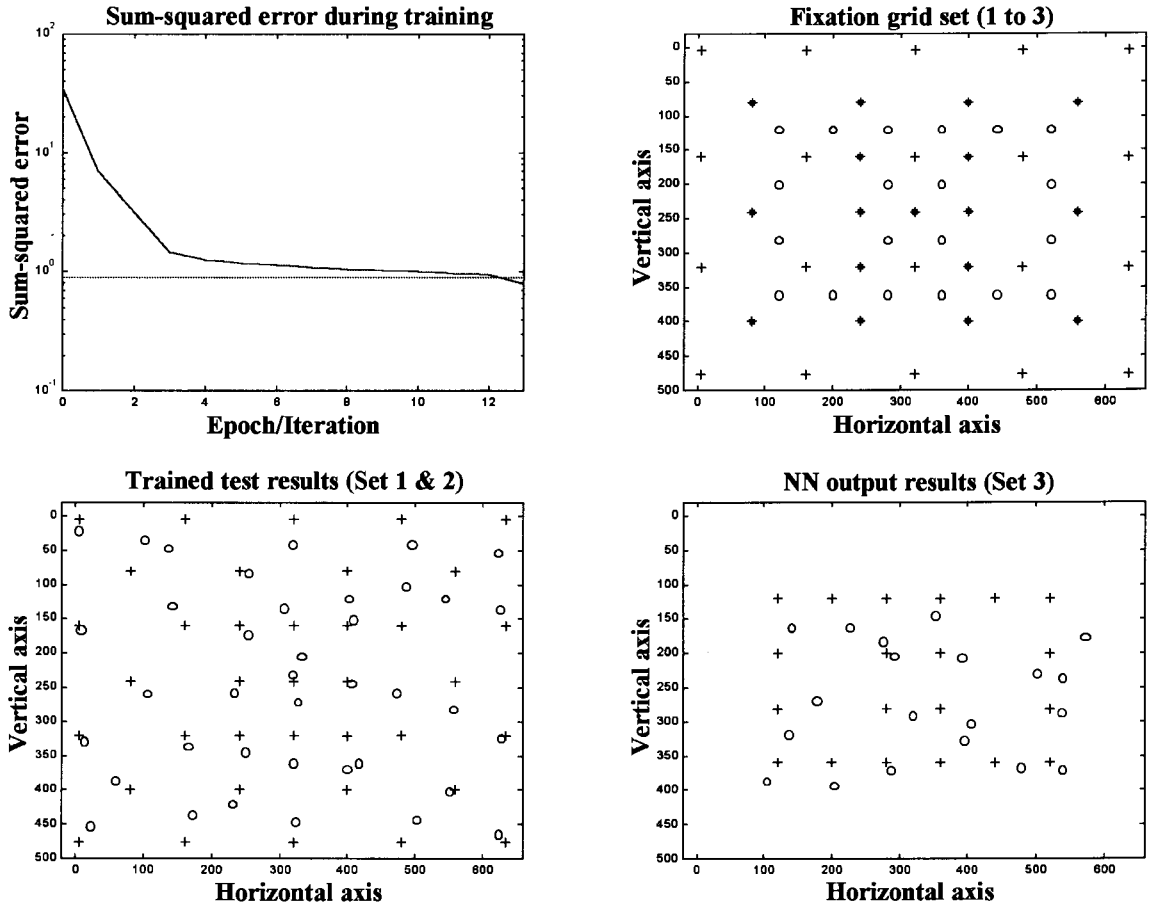

Figure 5.16 NN training results and calibration sets.

In the top left of figure 5.16 the sum-squared error values are indicated per epoch. Only the first and second set are used in training the NN. The third set is used to evaluate the NN performance.

In the bottom left of figure 5.16 the training results on the first two sets are imposed as circles on the calibration positions. In this particular training instance 8 hidden nodes were used with an error threshold of 0.9. The test results are shown in the bottom right.

In this test procedure the NN mapped the eye features to pixel position. The fixation points are shown with a 640x480 resolution and viewed at a 25° horizontal field of view. Therefore the mapped error deviation must be less than 25 pixels for one-degree accuracy. The first line in table 5.1 summarizes the training results of figure 5.16. The trained pixel error columns give the mean pixel error, obtained after training with set 1 and 2, in the horizontal and vertical direction respectively. In the next two columns show the mean pixel errors obtained with test set 3. A vast number of variations in the NN structure, parameters and data preparation have been tried to determine the optimum NN structure. The other rows listed in the table below serve as an illustration of some of effects found in the development phase.

Table 5.1 NN Prototype Platform Training Results

NN	Hidden Nodes	Error Threshold	Trained Pixel Error		Tested Pixel Error		Epoch Iterations
			$\bar{\varepsilon}_{px}$	$\bar{\varepsilon}_{py}$	$\bar{\varepsilon}_{px}$	$\bar{\varepsilon}_{py}$	
1	8	0.9	9.9	27.8	26.1	31.9	13
2	8	0.5	8.6	18.2	24.8	48.3	38
3	12	0.5	8.1	16.7	35.8	54.3	27
4	6	0.9	16.1	25.9	25.7	26.7	21
5	6	0.9	11.9	27.1	30.1	37.6	9

In all cases one can see that the vertical eye movements are less accurate, which indicates that the feature extraction process's vertical sensitivity needs to be improved. This can be anticipated, because the main artifacts in the centroid extraction technique are caused by the eyelashes, which induce large variations in the vertical direction. Note that the third set's points do not include any of the training points and enables one to check overfitting and underfitting conditions.

One of the most important factors observed during training is that the network must not be trained too excessively, otherwise the generalization is drastically reduced. In the second row of table 5.1 the sum-squared error threshold has been lowered, which resulted in improved pixel errors during training. However, the mean pixel errors for test set three increased. By lowering of the error threshold the network must concentrate on the outliers or fit the other points more stringently, which in either case adversely effects the generalization property of the network.

Merely increasing the number of hidden nodes, decreases the training error, but leads to overfitting conditions and vice versa conditions when lowering the number of nodes. Row three in the table shows typical results obtained with increasing the number of nodes. The best results are listed in row four, with a six-node network. In the next row however, the same NN structure shows contrary results. This is typical of underfitting conditions. In these circumstances the choice of initial weight conditions also plays a major role.

From the tests we learned that an eight-to-ten hidden node layer, with a generous error threshold setting offered the best results. However, performance varied drastically between subjects and it was difficult to find a single network structure that exhibited reproducible results for various subjects. The robustness constraint could therefore not be satisfied at the prototype stage, using the initial feature extraction routines and NN implementation.

The results looked promising though and the NN concept seemed viable, but some work still had to be done to improve robustness and reproducibility.

5.3.2 Bisection NN implementation

With the bisection feature extraction method the NN model has to map the PPD to the gaze position within the visual domain, which is determined via the servo control commands.

5.3.2.1 Topology

Only two input neurons are required for the horizontal and vertical PPD differences. Two output neurons supply the servo commands that control the projection position of the foveal image. Figure 5.17 shows the topology of the NN implementation.

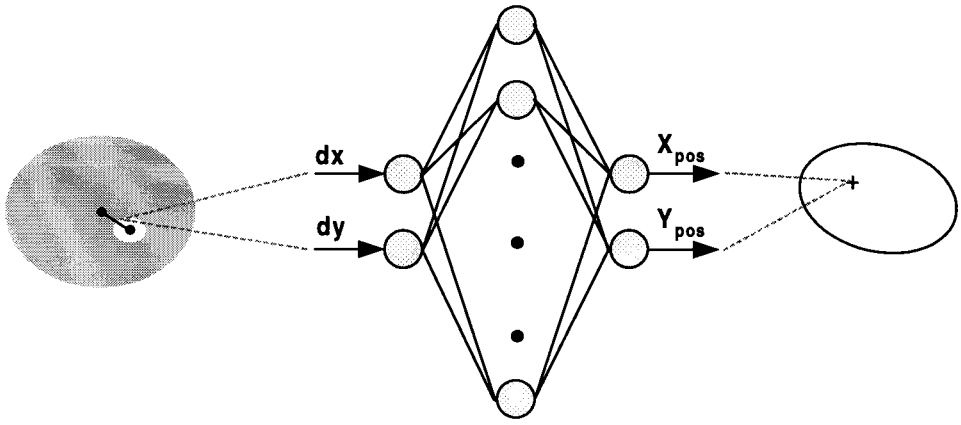

Figure 5.17 NN structure for the bisection implementation.

The architecture and learning procedure of the network is similar to the prototype platform and is implemented with ten neurons in the hidden layer.

5.3.2.2 Calibration & Results

The calibration process is similar to the procedure described in the prototype platform NN implementation. Two calibration sets, each containing 25 fixation points, are used. The first set is scanned horizontally and the second set vertically. Instead of screen coordinates, the sets now contain mirror coordinates, or the servo commands required to project the foveal image at a certain point in the visual domain.

In the prototype implementation a flashing cross is shown on the PC monitor and moved to the next calibration position with each button press. In the SFPD system the flashing cross is always centered in the CRT display and is projected within the visual domain by means of mirror movement; controlled via the servos. Figure 5.18 depicts this setup.

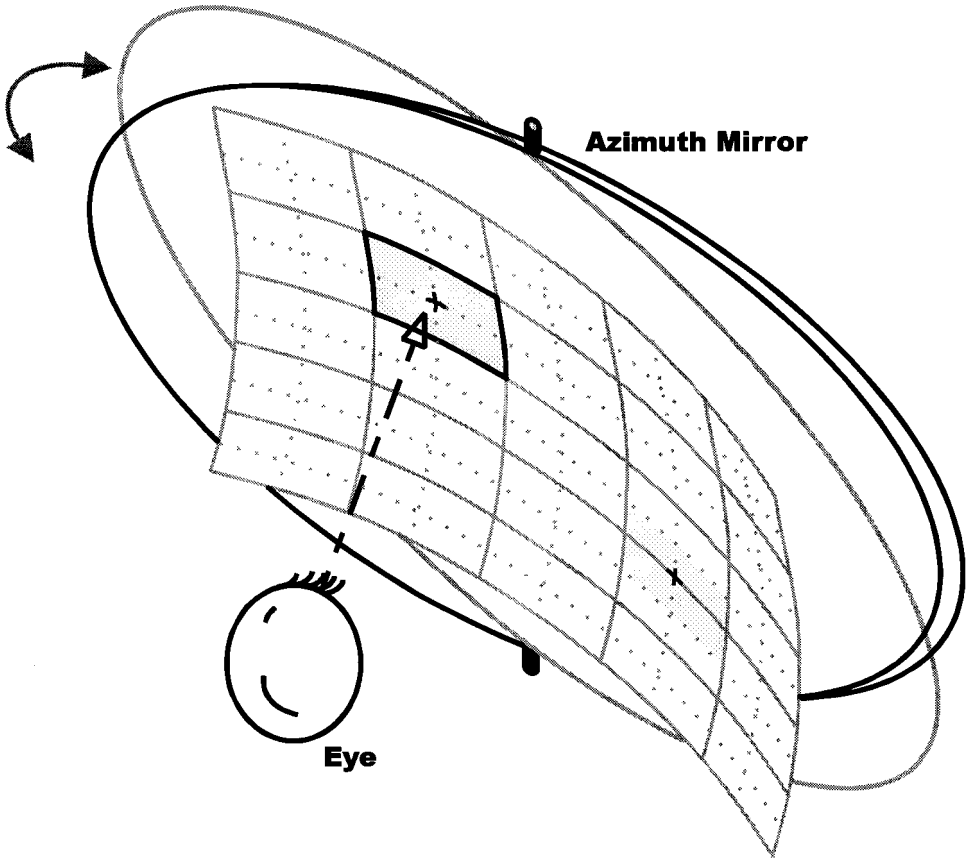

Figure 5.18 Calibration procedure for the SFPD NN implementation.

Figure 5.18 neatly demonstrates the second approach discussed in section 1. We know that the NF has a ±10° field of view and that the mirror movement is calibrated within these limits. The servo commands are 12-bit RS-422 serial commands. By simply positioning the mirrors at uniformly distributed positions in the visual domain and recording the extracted eye features from the bisection method, the NN can be trained to do the mapping between the parameters. Via this approach the transformation process from features to gaze position is embedded within the network, without any need for detailed system equation formulations. This is an elegant solution with the additional advantage that any non-linear characteristics within the mapping process can be dealt with effectively.

More detail and results of the calibration procedure and NN performance results are given in chapter six. In order to illustrate and compare the increased performance of the bisection method NN implementation against the prototype NN implementation the center fixations in the fixation sets were excluded in the training process and used as test points. The calibration points are shown on the left of figure 5.19, with the circles indicating the center test points.

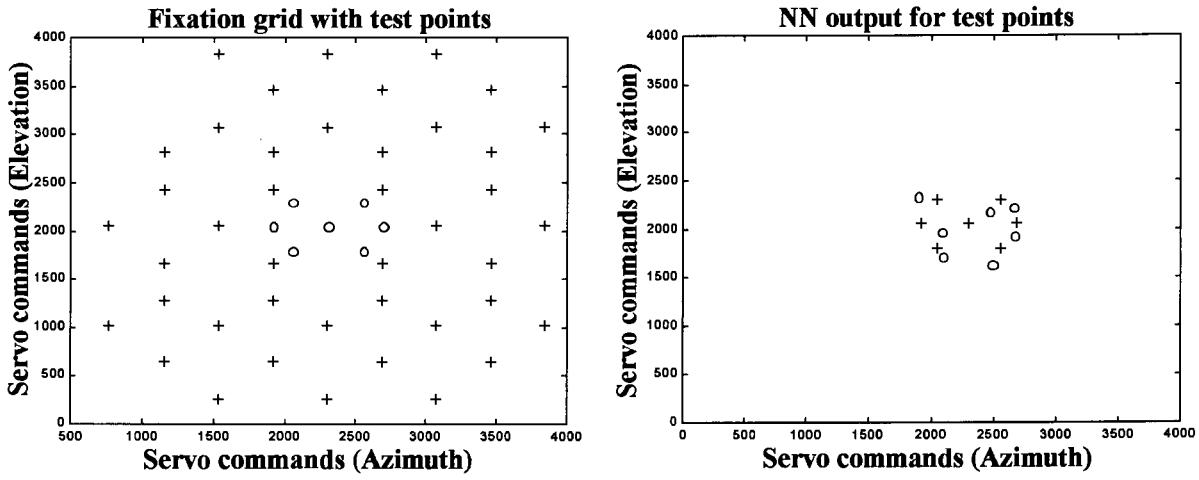

Figure 5.19 NN calibration points and training results.

Once trained, the excluded points are used to evaluate the NN performance. The result is shown on the right of figure 5.19 and listed in table 5.2. In this setup we have 20 degrees covered with a 12-bit resolution, or 200 units per degree. The mean tested error listed in table 5.2 is well below this threshold and is consistently so for all subjects.

Table 5.2 NN Training Results *(Bisection method)*

Hidden Nodes	Error Threshold	Trained Error		Tested Error		Epoch Iterations
		$\bar{\varepsilon}_{px}$	$\bar{\varepsilon}_{py}$	$\bar{\varepsilon}_{px}$	$\bar{\varepsilon}_{py}$	
10	5	45.19	93.77	104.41	100.96	28

These results verify the use of the NN concept in the mapping functional block, proposed in section one. With the use of the more robust bisection feature extraction algorithm, the NN's robustness and reproducibility increased drastically. With the knowledge obtained in the

prototype implementation the optimal NN configuration for the new topology was determined quickly and the embedded network needed only minor modifications. The results prove the effective use of a NN structure in the eye directed system.

5.3.2.3 Post-filtering

System noise contained in the input caused the NN to continuously make small adjustments in the output, which resulted in small mirror jitter movements centered at the gaze position. If the images are fused these jittery movements do not disturb the subject, as an updated NF foveal image is always available at the gaze position within the fusion boundary constraints. In the test conditions where one view the NF crosshair image against a blank WF, or in the presence of a definite illumination difference between the WF and NF images in fusion these small movements are quite distracting. Some post-filtering steps have therefore been introduced to eliminate these noise artifacts.

With 1° RMS specified for fusion the NF image can be stationary until the gaze position exceeds the gaze position boundary, implying the use of a dead band filter. Thus the output can actually exhibit a small lag or dampening effect which inhibits the jittery motion; the exact opposite of the above scenario. An optimal dead band filter implementation would require a threshold setting just above the system noise level. In our application this threshold was determined via an iterative process by progressively increasing the threshold level until the jittery motion was filtered out. If the threshold is too high one gets a "sticky" NF image where eye movements outside the fusion window are required to move the foveal image to its new position. *(This effect can be quite useful in a drag-and-drop scenario where only large eye movements position the NF image/marker and allows small eye movements within the gaze area without removing the foveal spot or marker from the main interest point.)*

The post-filtering implementation continuously averages two successive NN outputs and only adjusts the mirror position if the difference in gaze position exceeds a threshold level of 0.5 degrees and is less than 7.5 degrees between two frames. The upper threshold prohibits spurious noise spikes. With post-filtering the eye tracker exhibits very smooth gaze projection characteristics. The post-filtering technique also effectively removes eye blink artifacts via the upper threshold setting.

Chapter 6

Evaluation Procedure and Results

In this chapter the test procedure used in the evaluation of the eye-directed controller's performance is described. In the first section we discuss the final calibration and NN training procedure. Sixteen subjects were used in the evaluation tests. The results of these tests are presented in the second section.

The evaluation procedure for the eye-directed system forms an integral part of the total SFPD system. Evaluation of the final system accuracy was therefore not the only objective, but it was also important to determine the following:

- How "natural" subjects found the interface.
- Characterization of the learning process involved.
- Repeatability and robustness of the algorithms.
- Can subjects master the process, even in this bare concept phase?
- The dynamic interaction performance within the total system.

The ergonomic evaluation of the total system and the viability of the SFPD system were some of the objectives of Viljoen's work and more information can be found in his thesis in [10]. In evaluating the performance of eye controllers/measurements, we concur with the following statement in Sheena and Borah's work [23]:

> "Note that there is a difference between precision and accuracy of eye movement measurement. Accuracy indicates how closely the measurement corresponds to the true eye position, while precision is a measure of the reproducibility of the measurement. In some applications, like reading, accuracy is important, while in others, like the study of eyeball dynamics, only precision is important."

As mentioned in previous chapters, the emphasis of the design described in this dissertation is definitely on reproducibility, alias precision, rather than accuracy.

6.1 Evaluation Procedure

During the evaluation process each subject is first briefed on the test procedure. At this point each subject is familiar with the system and had previous trials in the NN calibration procedure.

The main steps in the evaluation procedure are:

- Recording of the calibration fixation points.
- Verification and preparation of the recorded calibration data.
- NN training on recorded data.
- Verification of NN performance. *(Store trained NN weights for each subject.)*
- Perform evaluation test.

The steps prior to evaluation are repeated until successfully verified. Each of these stages is discussed in the following sections.

6.1.1 Recording and Verification of Calibration Data

As mentioned in chapter 5, the calibration process consists of a number of sequential target points on which the subject must fixate. A grid pattern with 25 fixation locations is used. Two sets are used: in the first set the points are mainly presented in a horizontal direction, while vertical eye movements are required in the second set. Remember that the best training results were achieved when the targets were presented in a straight-line manner, rather than in a random-like fashion. Figure 6.1 shows the uniformly scattered fixation points in the visual domain.

In order to provide better sensitivity or accuracy in the center of the visual field, more points are located in this area. A slight protrusion of the mechanical mechanism in the top right-hand corner of the visual field sometimes obscures the fixation point in this area. Therefore the grid pattern is not symmetrical, as the fixation point in this region has been removed. The target or fixation point is simply a flashing crosshair that is shown in the center of a blank,

gray-white NF display that is positioned in the visual space via the servoed projection mirrors. A blank, gray-white WF display is presented to the left eye.

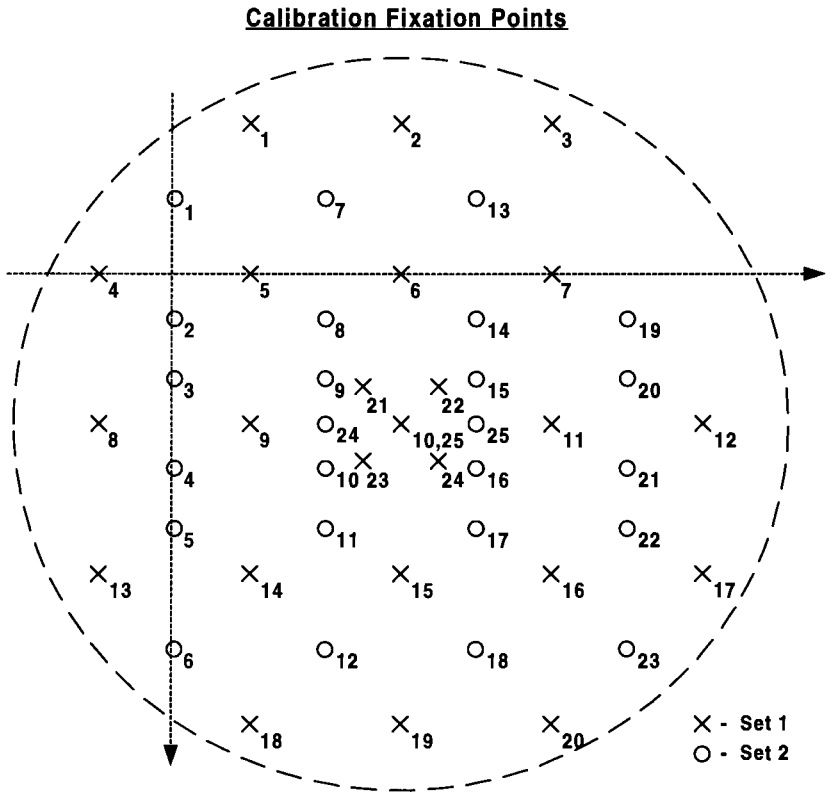

Figure 6.1 Calibration Fixation Grid Positions within the Visual Domain

The observer must fixate on the target and press a button when ready. On pressing the button the target will briefly appear and then disappear for a predetermined period, before it is presented at the next location. At each fixation 10 samples of the PPD features are taken within subsequent frames. (*250 samples per set*)

In the first calibration attempts the fixation point immediately jumped to its next grid position with the press of the mouse button. With no delay period set between subsequent fixation positions some subjects anticipated the next target position and immediately shifted their gaze to the next fixation point position on pressing the button. Using a sufficient delay period between target points solved this problem.

It was mentioned in chapter 2 that a person cannot hold his eyes perfectly still during fixation and that tiny eye movements occur several times a second. It is also not possible to move

your eyes in a smooth, continuous fashion along a straight line on a blank screen. In order to perform smooth, continuous eye movements a stimulus is needed that moves along a straight line and thus induces smooth pursuit movements. Similarly it is impossible for a person to lock his fixation at a test point in the NF display. Therefore, to aid the fixation process during calibration, a slow blinking target is used.

The following steps are followed during the calibration process:

- The person must be comfortable in order to keep his posture as still as possible during each set. *(Minimize any head-movement during calibration.)*
- If it is a person's first calibration, the first trial is only used to familiarize the person with the environment.
- It is important that the subject's eye is centered in the video display to ensure optimal tracking of the eye and to ensure that the person would be able to see the fixation points shown on the periphery.
- The subject must concentrate not to blink his eye on pressing the button.

6.1.1.1 Calibration Verification Process

At each fixation 10 samples are taken to enable an average eye position due to the involuntary, small eye movements. Once the calibration process is finished the samples are downloaded to the PC for verification. Invalid samples due to eyeblinks or incorrect centroid calculations are automatically detected and removed. If the characteristics of the recordings are inexact the calibration process is repeated. In most instances the calibration results improve drastically up to the third calibration trial, which confirms the anticipated learning curve requirement. Once the calibration points have been verified they can be passed on to the NN training phase.

Each sample in the calibration process contains the dx_{PPD} and dy_{PPD} features of the PPD. The mean and standard deviation are calculated for each fixation sample group that consists of ten sample points. Out of each group the four, best samples are selected and saved for training purposes. The selection process consists of an iteration procedure that is applied separately on the vertical and horizontal PPD samples in each group.

In this process the sample with the largest deviation from the mean of the remaining samples is recursively removed, until only four samples remain. Once selected, the variance for the reduced sample set is calculated and compared with the initial variance. The final variance parameter for each group should be less than 1. Finally the verified calibration data is imposed on the initial data set for visual inspection.

Figure 6.2 depicts the horizontal and vertical movement patterns between successive fixation points for the first and second set respectively. *(Refer to figure 6.1 for the fixation grid.)* It clearly depicts the mainly horizontal movement between fixations in the first set and vice versa for the second set. One should anticipate the same stepwise increment pattern and sweep pattern tendency in the recorded PPD data set.

The effect of the verification process can be seen in figure 6.3 where the vertical raw sample data of a recording is plotted against the verified data set. Here one can clearly see how the verification process removes the artifacts contained in the raw data. Note the similar pattern tendency between the sampled data and the original calibration data shown at the bottom of figure 6.2.

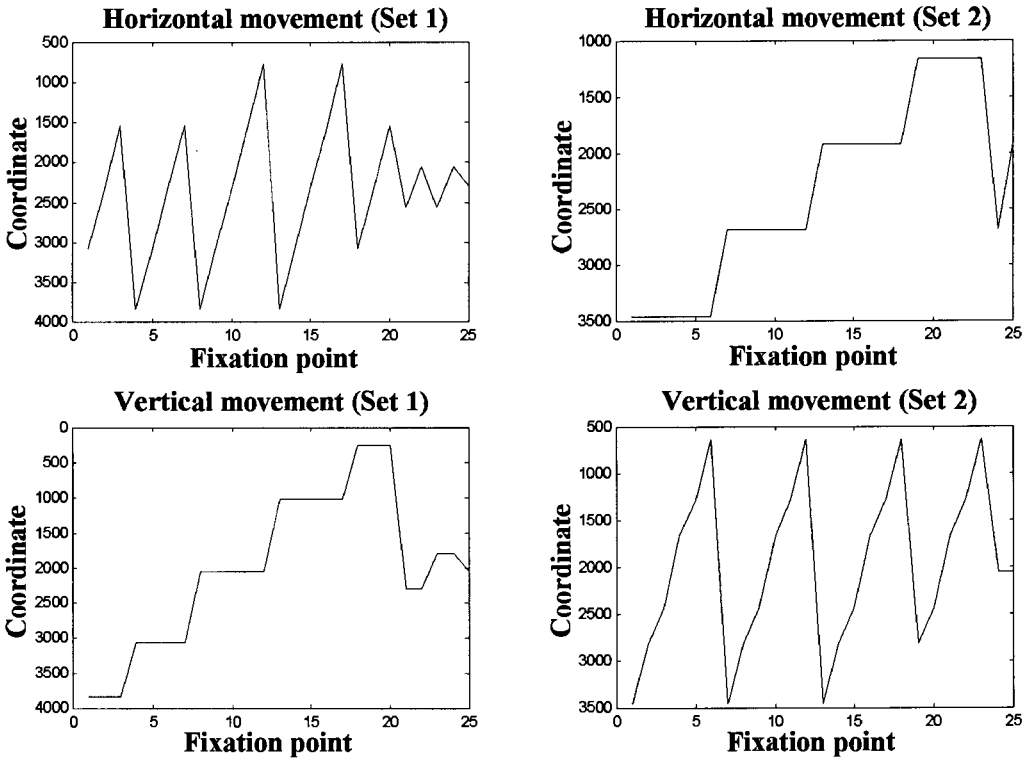

Figure 6.2 Calibration grid movement patterns between successive fixations.

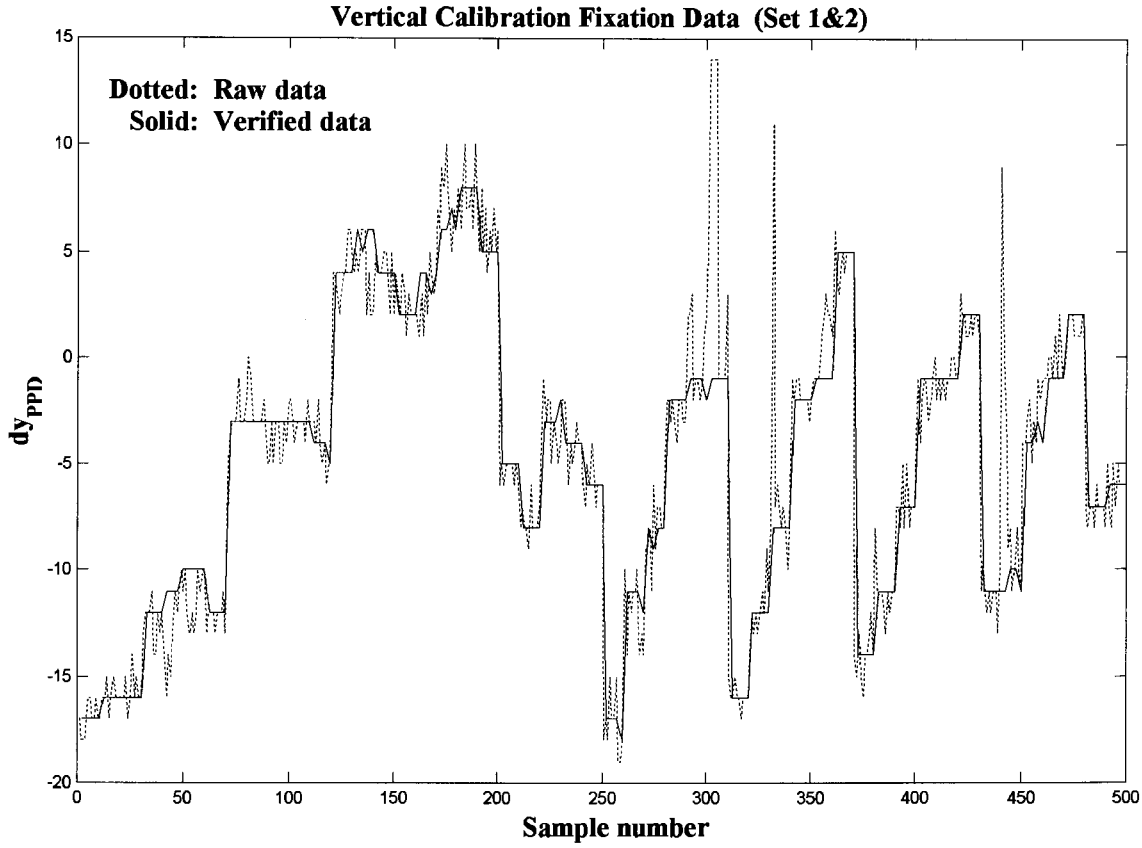

Figure 6.3 Vertical verified calibration data superimposed on raw data.

Typical verification results obtained during the final evaluation are shown in figure 6.4 and figure 6.5; presenting the results obtained with subject 5. Figure 6.4 depicts the comparison between the standard deviation results for the raw and verified data, while figure 6.5 shows the horizontal and vertical filtered data imposed on the raw data. In this particular calibration the standard deviation of the filtered data has drastically improved, while the filtered data in figure 6.5 has well defined step and sweep regions. These criteria indicate that the calibration process has been successful and can be used for NN training parameters. It is important that some variance remains in the data in order to allow the NN training process to create a smooth mapping surface that would inhibit erratic behavior during involuntary, small eye movements.

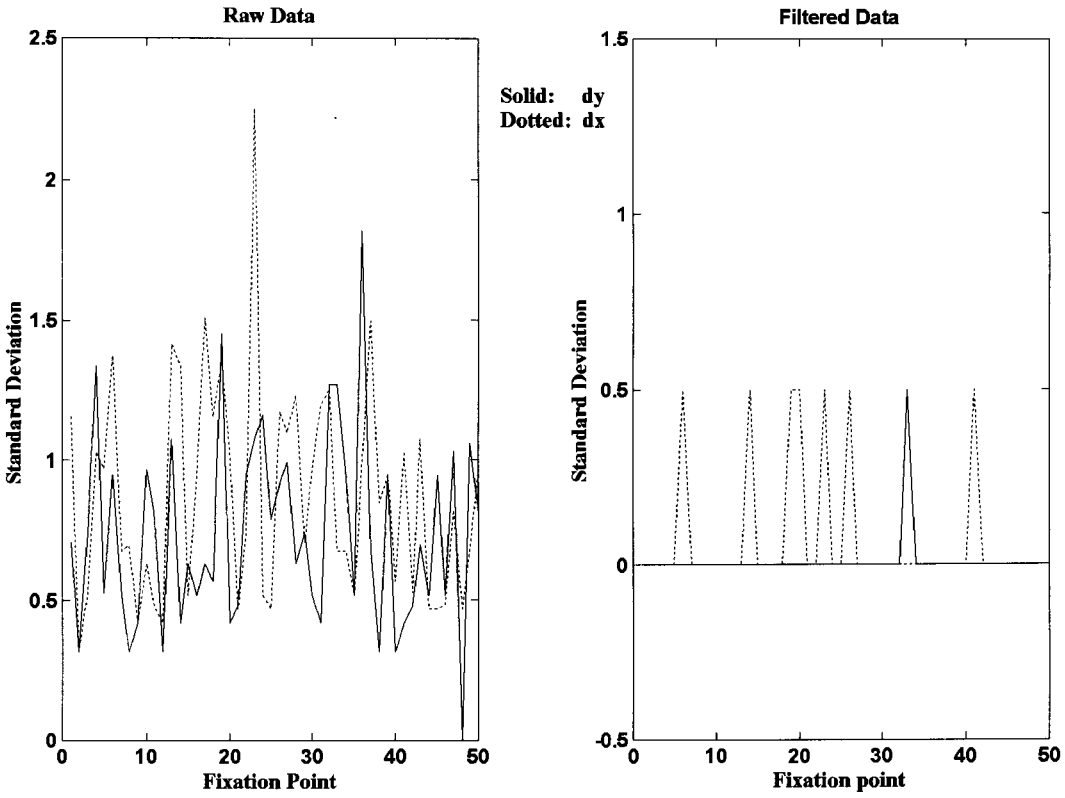

Figure 6.4 Raw and filtered standard deviation comparison for calibration data.

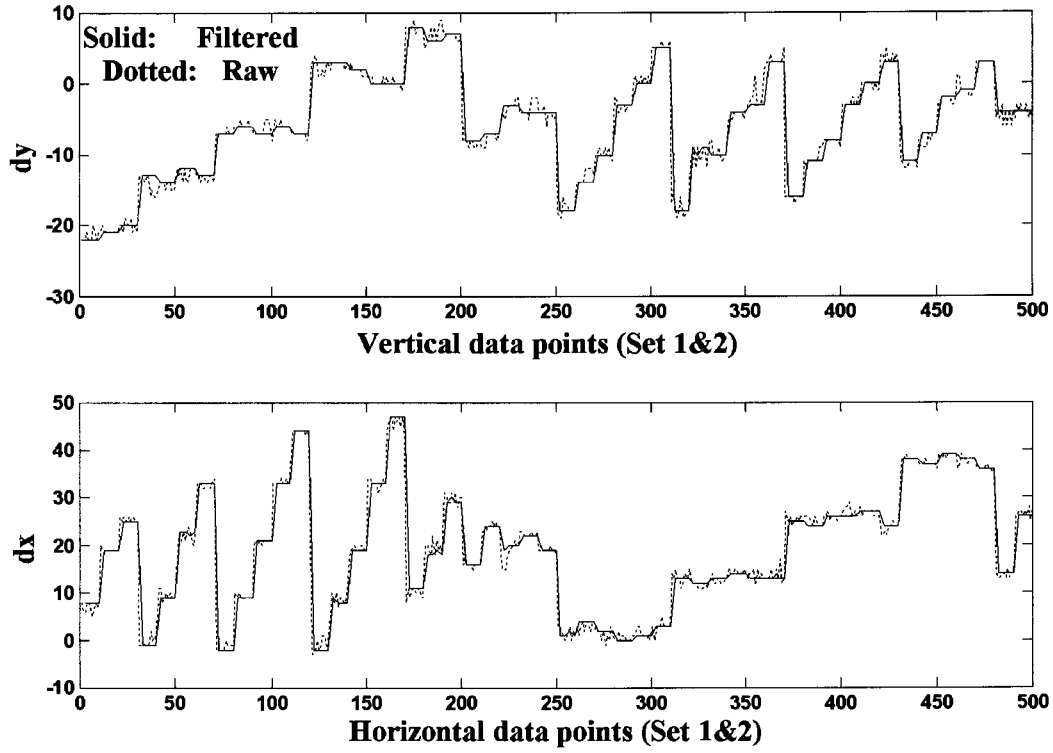

Figure 6.5 Typical filtered calibration data results.

Previously, without the verification process, artifacts in the samples were not detected and had a tremendous negative effect on the NN training process. During training a minimum cost function error must be reached. As the artifact values are completely different from the other sample points within the group, it prohibits a smooth map surface and requires more hidden nodes to decrease the cost function. If one simply allows for longer training or more iterations the process will "concentrate" on the irregularity in order to reduce the cost error, which would worsen erratic response behavior in this region.

A simpler method for verification purposes would simply be to use the average value of the sampled group, but in some cases the artifacts would cause a biased offset error in that region.

6.1.2 Neural Network Training Phase

Once the calibration fixation data has been verified, the NN can be trained on the processed input data, as described in chapter five. Throughout the preliminary tests in the evaluation of the NN performance, valuable information was gained on the NN training characteristics and a set of guidelines could be drawn up to evaluate the success of a training run. Without the proper verification process, as described in the previous section, it would have been very difficult to formulate any performance criterion on the training results.

If a particular training session is unsuccessful it can be repeated to ensure that it is not in response of the initial weight setup. However, an unsuccessful training session due to improper initial weights did not once occur and always was a result of some artifacts present in the training data set. As mentioned in previous sections, the calibration results improve drastically during the initial calibration trials; this improvement is clearly reflected in the success factor of the NN training procedure.

The optimal NN structure and fixed training parameters have been described in chapter 5. The main parameters or characteristics used to formulate the NN training guidelines are the following:

- The sum squared error threshold setting.
- The convergence characteristics of the training process.
- The number of iterations or epochs.

Without the verification of the input data, and a fixed NN structure one has to go through an iteration process, continuously changing parameters and comparing results between subjects. In these circumstances it was found that the best criterion for halting the training process is not a fixed error threshold setting, but a dynamic threshold that stopped training as soon as the error difference between subsequent epochs falls below a fixed level.

A sum squared error value of 10 was used in training and the error had to converge within 10 epoch iterations. Whenever the training process did not meet these criteria the calibration procedure was repeated.

6.1.3 Evaluation Test Procedure

In order to evaluate the eye-controller performance, fixation test points are displayed at known positions within the WF display and the operator is asked to fixate on each point. At each test-point the servo output commands are recorded and compared to the true position. Where the WF display was inactive during the calibration process, it is now used as reference in the evaluation; providing us with an inverse process for analysis purposes.

Once the NN training has been completed, the operator's NN weights are downloaded to the VFG board and evaluation can commence. A two-degree grid background is shown in the WF display and a crosshair in the NF display, while the operator's eye is being tracked. The operator is asked to fixate at the center of the grid and then to move his gaze along the vertical and horizontal axis respectively, whereafter he can fixate at random at any point within the scene. During each transition he must give feedback about the movement of the crosshair with respect to the background grid. This is similar to moving a cursor on your monitor display, except that one moves the cursor by simply fixating at the point of interest. The background grid with an imposed crosshair near the center is shown in figure 6.6.

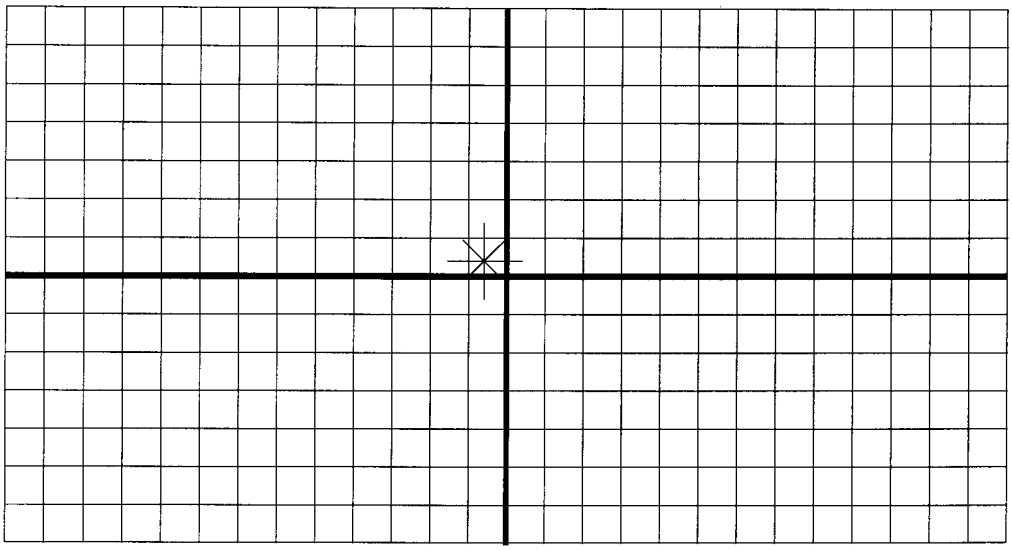

Figure 6.6 Two-degree background grid with imposed NF crosshair.

These actions fulfil the final calibration checks and indicate whether there is any skewness, inactive areas or other artifacts present in the track process. If any problems exist, the calibration and training sequences are repeated. It is important to tell the person that he must concentrate on the background or point of interest and then to note the crosshair position, otherwise he keeps trying to follow or direct his own gaze position.

In the next step the WF and NF displays must be aligned. The crosshair in the NF display is projected in the center of visual domain and the operator is asked to direct the movement of the grid background in the WF until the center of the grid is aligned or fused with the crosshair. At this point the brightness of the NF display is turned down in order to eliminate any disturbances by the mirror projection. Finally the operator is presented with nine sequential flashing test points in the WF display. At each test point the operator is asked to blink his eyes, fixate on the test point and press the mouse button when ready. He must keep his fixation on the point until a beep sound indicates the end of the recording, when the next test point is shown. At each test point 100 data points are sampled for both the azimuth and elevation command channel. The test point is presented as a flashing x that is positioned on the evaluation grid. The evaluation grid and test point locations are depicted in figure 6.7.

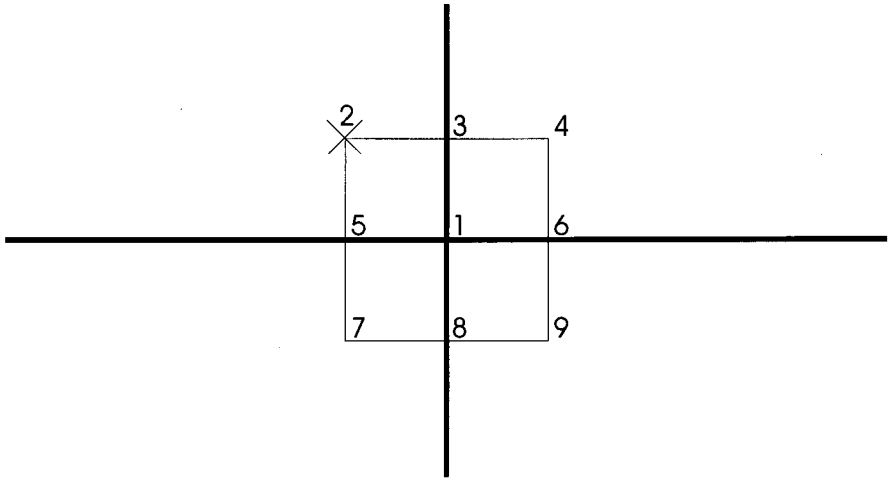

Figure 6.7 Evaluation grid with fixation test point locations.

The recorded data can now be compared with the known projection co-ordinates and enable us to evaluate the precision of the eye-directed controller.

The steps in the test procedure can be summarised as follows:

- Calibrate fixation points.
- Verify calibration process and repeat if necessary.
- Prepare the calibration data for NN training process.
- Train the NN within guidelines and save the weight parameters.
- Download the subject's weight parameters and check initial performance.
- Do fixation accuracy performance test.
- Evaluate and analyze results.
- Continue with SFPD evaluation.

Subsystem results have been shown and discussed within the chapters where the design and refinement of the relevant function have been addressed. In this section we present the results obtained in the final evaluation procedure.

Figure 6.8 and 6.9 on the next page depict the typical results obtained for an individual. In the first figure the encircled plus signs indicate the nine test points within the visual domain. The recorded data is plotted in crosses. The dotted line connects the data to indicate the movement between successive fixations. Intuitively one would expect a cloud of crosses near each fixation test point. However, as discussed in chapter 5, the dead band filter suppresses unwanted mirror movement near a current fixation and will only update with movement exceeding 0.5°. Therefore we only see isolated crosses plotted near fixation points; indicating the effectiveness of the dead band filter. The transition between successive points for this case is plotted separately for the azimuth and elevation channels in the second figure. Note that the error with regard to the reference signal seldom exceeds a value of 200 *(1 deg)*.

In order to evaluate the accuracy of the eye-directed controller for this case, the mean azimuth and elevation values recorded at each test point are plotted in figure 6.10. The dotted boundary squares depict the one-degree accuracy regions at each point, with the mean data position indicated with circles. This particular example clearly shows that the recorded data is well within the one-degree RMS specification. Figure 6.11 depicts a similar graph for the data of all sixteen subjects, with 1.5° demarcated areas indicated.

The results obtained for the 16 subjects are summarised in table 6.1. The second column presents the average error value over the 900 fixation samples for each test subject. The standard deviation values for the recordings are shown in the third column. It is evident from these figures and the table summary that the eye-directed controller achieved its objective with an accuracy well within the 1 deg specification.

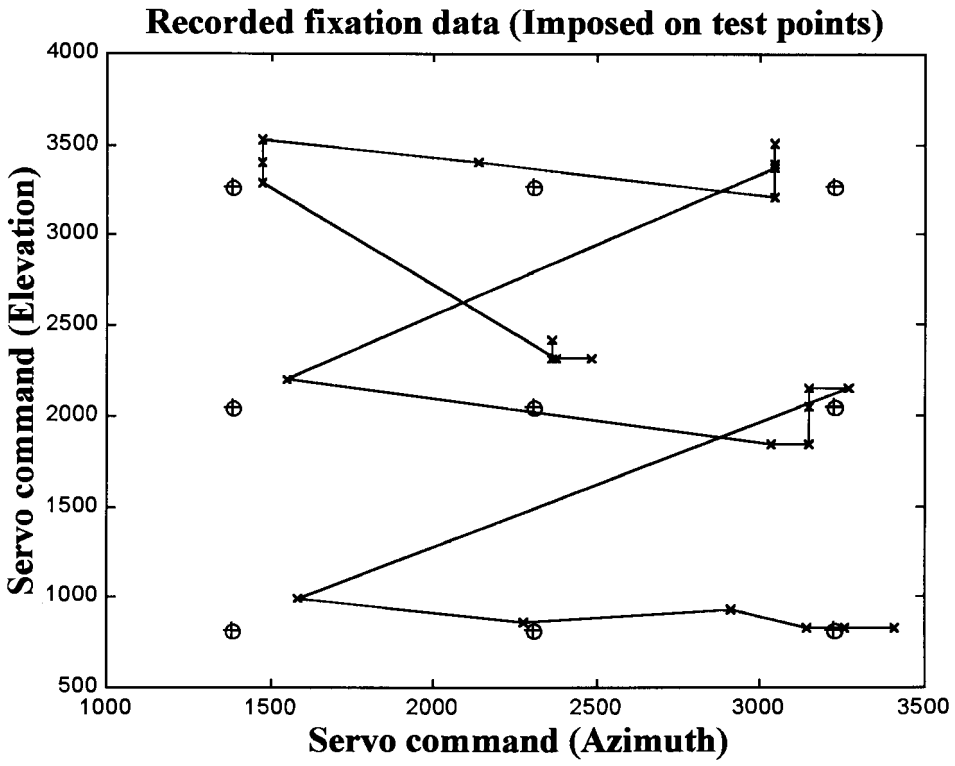

Figure 6.8 Fixation evaluation data imposed on test points.

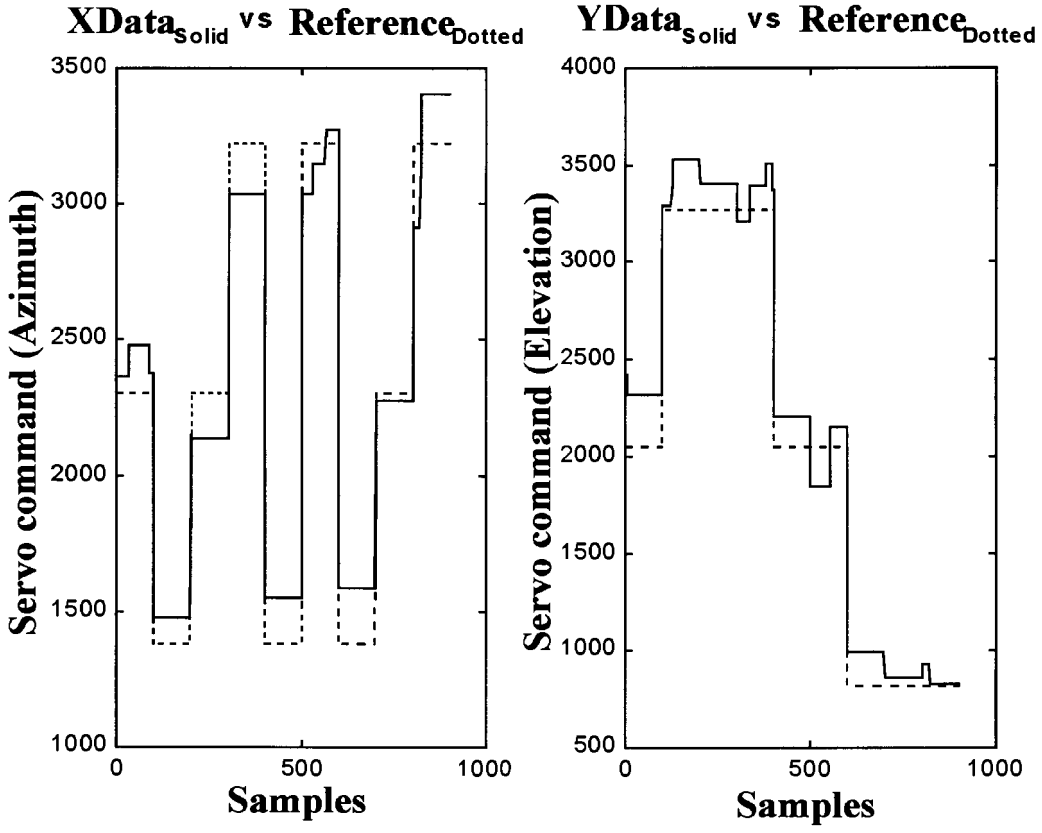

Figure 6.9 Sampled test data vs. reference positions.

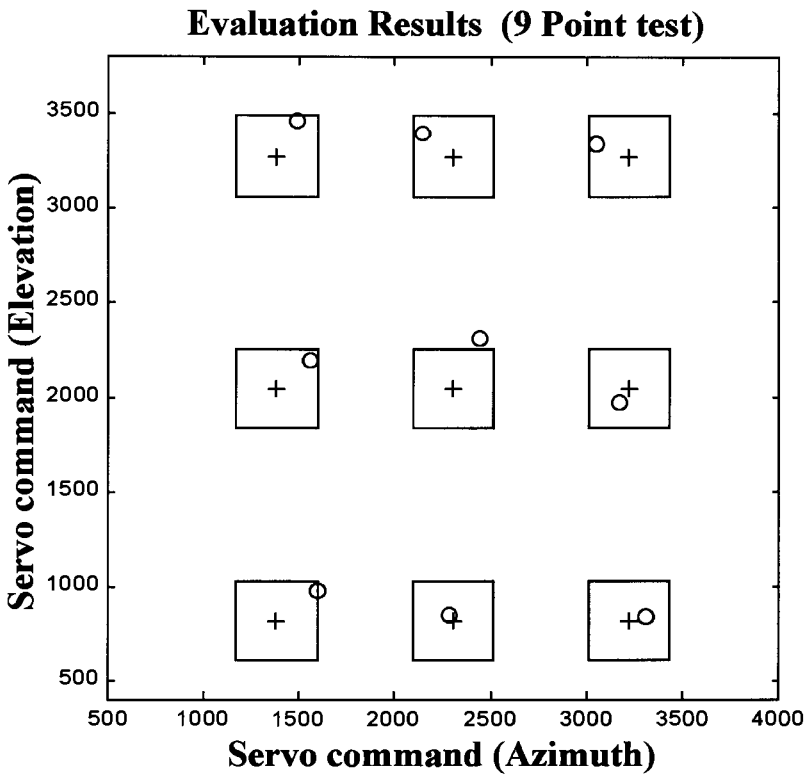

Figure 6.10 Evaluation results with 1° boundary boxes.

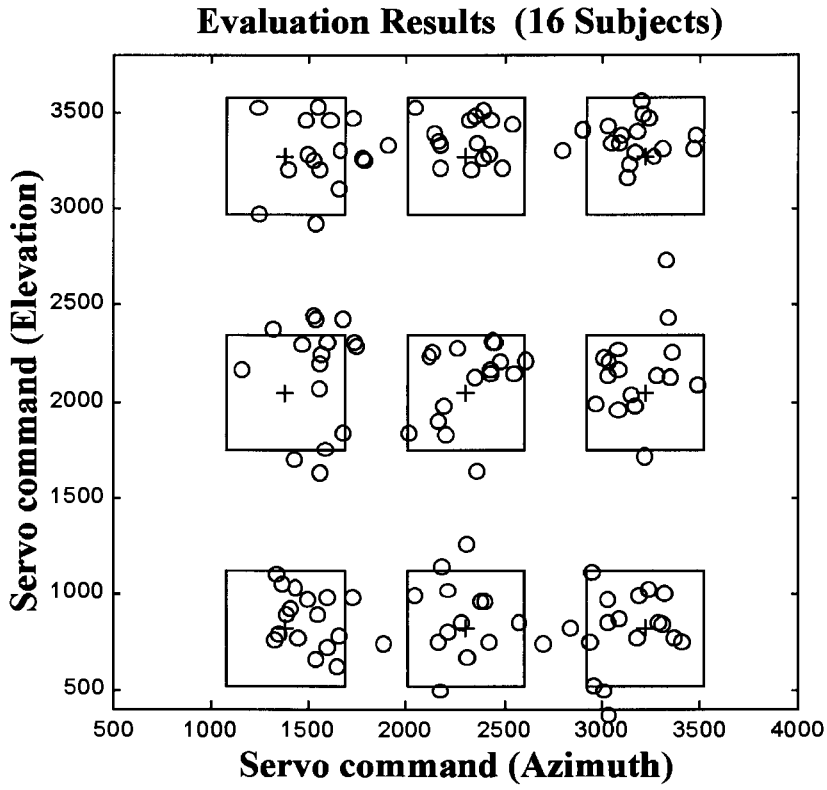

Figure 6.11 Combined evaluation results with 1.5° demarcated areas.

Table 6.1	**Summary of evaluation procedure results.**	
Subject	**Gaze Error Mean (Degrees)**	**Gaze Error Standard Deviation (Degrees)**
1	0.55	0.87
2	-0.59	0.76
3	-0.76	0.98
4	-0.84	0.93
5	0.70	0.94
6	-0.57	0.89
7	-0.15	0.84
8	-0.76	0.66
9	-0.40	0.82
10	0.04	0.77
11	-0.61	0.59
12	-0.33	0.95
13	-0.09	0.85
14	-0.34	0.62
15	0.28	0.76
16	-0.16	0.88
Avg.	-0.25	0.82

Chapter 7
Discussion & Conclusion

7.1 Discussion

The first step in the development process was the development of the PPD image processing feature extraction routines. Once the initial feature extraction algorithm performed fairly well in the prototype configuration, all attention was devoted to the development of the NN implementation. With the SFPD setup the focus again fell on the NN design before it was realized that the feature extraction process needed attention. In the end the bulk of the time spent on the design was allotted to the feature extraction process. Although some time was wasted in refining the NN, valuable information was gathered in this process. Through recursive optimization of the NN topology and training parameters, training characteristics specific to the implementation could be identified and used to verify the training process in the latter stages.

In [15] Wagner concludes: "the next stage would be to implement the algorithm in hardware, with real-time capabilities." He suggests that in order to reduce computational cost some intelligent registration strategies would be needed. This is exactly what has been achieved in this dissertation, although it has not been the primary objective. A novel image processing extraction algorithm has been proposed and experimentally verified, that is invariant to partial occlusions and has been implemented in real-time.

One of the most severe limitations in the current eye tracker and the cause for many of the anomalies is the IR illumination. The current IR LED setup is very sensitive and the LED must be quite bright to provide adequate illumination, especially with persons with deep set eye-sockets. The off-axis location of the LED inhibits the gaze angle and head movement tolerances within the helmet. A diffused illumination source must be used that provides adequate, uniform illumination of the eye and does not cause multiple reflections in spectacles. A small bright source on the optic axis can then be used for the purkinje reflection. This would eliminate multiple reflections, ensure a centered purkinje image that

would allow less sensitivity to head movements and ensure more accurate PPD extraction. In current commercial eye trackers holographic optics is used to address this problem.

Currently spectacles or hard contact lenses cannot be used in the system. The persons used in the tests all had 20/20 vision without any physiological eye disorders. People that have strong dominance in their left eye had problems with fusion. In these cases the NF projected image will have to be shown in their left eye.

An interesting factor that has been noted during calibration is that all subjects exhibited an initial learning curve in the calibration fixation procedure. The fixation ability of the subjects varied. In persons with repeatable accurate fixation capabilities excellent results were obtained. It would be interesting to see if these subjects also showed superior target detection capabilities. In this development the main focus has been on delivering an operational eye-directed controller. A study on the fixation recording characteristics and the relation between fixation patterns of different subjects has not been conducted.

Improved performance for the eye-directed controller can be obtained with better illumination and a better resolution CCD camera. *(non-interlaced 256x256 resolution per frame)* Once the feature extraction accuracy has been improved different NN structures can be evaluated for optimal performance.

7.2 Conclusion

In this dissertation it was experimentally demonstrated that a neural network controller can be successfully applied in a real-time eye-directed system. Especially within the SFPD configuration it provides an elegant solution in that the gaze position can be directly mapped to the projection co-ordinates without the need of any axis or co-ordinate transformations. By providing foveal and peripheral vision in a far-field teleoperator through the eye-directed SFPD, visual information is provided in a natural way and telepresence can be established.

Further proof of the eye-directed controller's performance is the fact that in Viljoen's SFPD tests the subjects successfully fused the images and could detect the targets in the test scenario. *(See [10] for more information.)*

In connection with the main assignments listed in the introduction we can conclude that:

- A novel image processing PPD feature extraction technique has been developed that exhibits a robust and reproducible real-time performance. *(Bisection technique)*
- A neural network controller has been successfully implemented that directly maps the gaze position parameters to the projection control commands.

More importantly the overall objective of developing an eye-slaved controller for a SFPD system has been achieved and was experimentally demonstrated.

7.2.1 Application and Research Possibilities

Within a more robust experimental setup the eye-controller can be used to study eye movements and to perform ophthalmological tests. Visual perception, gaze interest, target search and target detection strategies can be analysed.

The slaved foveal technique also provides an efficient information utilization system in that the background scene needs only be displayed in low resolution, requiring less bandwidth. Only the small foveal image needs to be displayed in high resolution. Virtual head-up

displays can utilize this fact and need only to update the scene in detail at the current gaze position. Instead of using a data-glove in virtual simulators the eye gaze position can be used for navigation and distinct eye-blinks can be used for selection purposes. Videoconference systems could perhaps use the foveal update technique, in tracking a person's gaze via a camera on the display. In this application only the window near the gaze position needs to be updated; the background can be updated every nth cycle.

References

1. Sheridan T B, "Defining our terms", *Presence*, Vol. 1, No. 2, pp. 272-274, Spring 1992

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

3. Tachi S, Tanie K, Komoriya K, Kaneko M, "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, pp. 245-254, 1984.

4. Kim W S, Matsunaga K, Stark L, "A Helmet Mounted Display for Telerobotics", *COMPCON Spring 1988*, 33rd IEEE Computer Society International Conference (no. 88CH2539-5), San Francisco, USA, pp. 543-547, 29 Feb – 3 March 1988

5. Hamit C G, *Virtual Reality and the Exploration of Cyberspace*, Sams Publishing, USA, 1993

6. Gerrison J F, "On the Network-Based Emulation of Human Visual Search", *Neural Networks*, Vol. 4, pp. 543-564, 1991

7. Young L et al, "Visual and Control Aspects of Saccadic Eye Movements", *NASA CR-564*, pp. 95-125, 1966

8. Goode P W, *Eye Directed View*, Automation Technology Branch, Langley Research Centre, NASA, Circa 1984.

9. Yamaguchi H, Tomono A, Kobayashi Y, "Proposal for a Large Visual Field Display Employing Eye Movement Tracking", *SPIE*, Vol. 1194, Optics, Illumination and Image Sensing for Machine Vision IV, 1989

10. Viljoen G T, A Comparative Study of Target Acquisition Performance between Teleoperated Sensors and Direct Human Vision, Doctoral thesis, University of Stellenbosch, South Africa, December 1999

11. Hubel D H, *Eye, Brain, and Vision*, Scientific American Library, USA, 1988

12. Marieb E N, *Human Anatomy and Physiology*, pp. 492-511, Benjamin Cummings, USA, 1990

13. Young L R, Sheena D, "Methods & Designs: Survey of eye movement recording methods", *Behaviour Research Methods & Instrumentation*, Vol. 7, No. 5, pp. 397-429, 1975.

14. Ogle K N, "Vision" in Pender H, McIlwain K (Editors), *Electrical Engineers' Handbook: Electric Communication and Electronics*, John Wiley & Sons, New York, 1950

15. Wagner R, *Image Processing Applied to the Tracking of Eye Movements*, Thesis, McGill University, Montreal, 1989

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

17. Zhou G, Ezumi K, Stark L W, "Efficiency of Searchpatterns", *Comput. Biol. Med.*, Vol. 23, No. 6, pp. 511-524, 1993

18. Noton D, Stark L, "Eye Movements and Visual Perception", *Scientific American*, Vol. 224, No. 6, pp. 34-43, 1971

19. White K P, Hutchinson T E, Carley J M, "Spatially Dynamic Calibration of an Eye-Tracking System", *IEEE Transactions on System, Man, and Cybernetics*, Vol 23, no.4, July/August 1993

20. Fisher F F, Monty R A, Senders J W (Editors), *Eye Movements: Cognition and Visual Perception*, Lawrence Erlbaum Associates Inc, USA, 1981

21. Carpenter R H S, "Oculomotor Procastination" in [20], pp. 237-246

22. Stark L, Ellis S R, "Scanpaths Revisited: Cognitive Models Direct Active Looking" in [20], pp193-226

23. Sheena D, Borah J, "Compensation for Some Second Order Effects to Improve Eye Position Measurements" in [20], pp. 257-268

24. Tole J R, Young L R, "Digital Filters for Saccade and Fixation Detection" in [20], pp. 247-256

25. Ditchburn R W, "Small Involuntary Eye-Movements: Solved and Unsolved Problems" in [20], pp. 227-235

26. Golding A P, *Search performance and peripheral visual acuity*, Report PERS-409, Human Science Research Council, Pretoria, 1986

27. Glenstrup A J, Engell-Nielsen T, *Eye Controlled Media: Present and Future State*, Thesis, University of Copenhagen, 1995

28. Barry W, Melvill J G, "Influence of Eye Lid Movement Upon Electro-oculographic Recording of Vertical Eye Movements", *Aerospace Medicine*, Vol. 36, No. 9, pp. 855-858, 1965

29. Hutchhinson T E, White K P, Martin W N, Reichert K C, Frey L A, "Human–Computer Interaction Using Eye-Gaze Input", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, No. 6, pp. 1527-1534, 1989

30. Pomplun M, Velichkovsky B, Ritter H, "An Artificial Network for High Precision Eye Movement Tracking", *KI-94: Advances in Artificial Intelligence – 18th German Annual Conference on Artificial Intelligence*, pp. 63-69, 1994

31. Cleveland N R, Doyle R N, "Eyegaze Computer System: How does it work? Who can use it?", *Closing the gap – 10th annual conference*, Minneapolis, Minnesota, Oct 1992

32. Yamada M, Fukuda T, Hirota M, "A New Eye Movement Analyzer: Auto Calibration and Wireless Transmission", *IEEE Engineering in Medicine & Biology Society 11th Annual International Conference*, 1989

33. Howard Demuth, Mark Beale, *Neural Network Toolbox User's Guide*, The Math Works Inc, 1994

34. Eberhart R C, Dobbins R W, *Neural Network PC Tools: A Practical Guide*, Academic Press, USA, 1990

35. Dufaux F, Moscheni F, "Motion Estimation Techniques for Digital TV: A Review and a New Contrinution", *Proceedings of the IEEE*, Vol. 83, No. 6, pp. 858-876, June 1995

36. Rubenstein M, Assessing Target Centering Algorithms for use in Near-Real-Time-Photogrammetry, Masters Thesis, University of Cape Town, 1990

37. Pratt W K, *Digital Image processing*, Second Edition, John Wiley & Sons, New York, 1991

38. Spiegel M R, *Mathematical Handbook of Formulas and Tables*, McGraw-Hill Inc., New York, 1968

39. Frecker R C, MacLean W J, "Linear Multivariate Regression Technique for Calibrating 2-Dimensional Corneal-reflection Eye Trackers", *CMBEC-15-CCGB*, Toronto, 1989

40. Ebenholtz S M, "Hysteresis Effect in the Vergence Control System: Perceptual Implications" in [20], p83-94

Appendix A
Formulas

1 Perimeter point algorithm

By using the "Equation of circle of radius R passing through origin" in [38] the equations needed for this process was determined as follows:

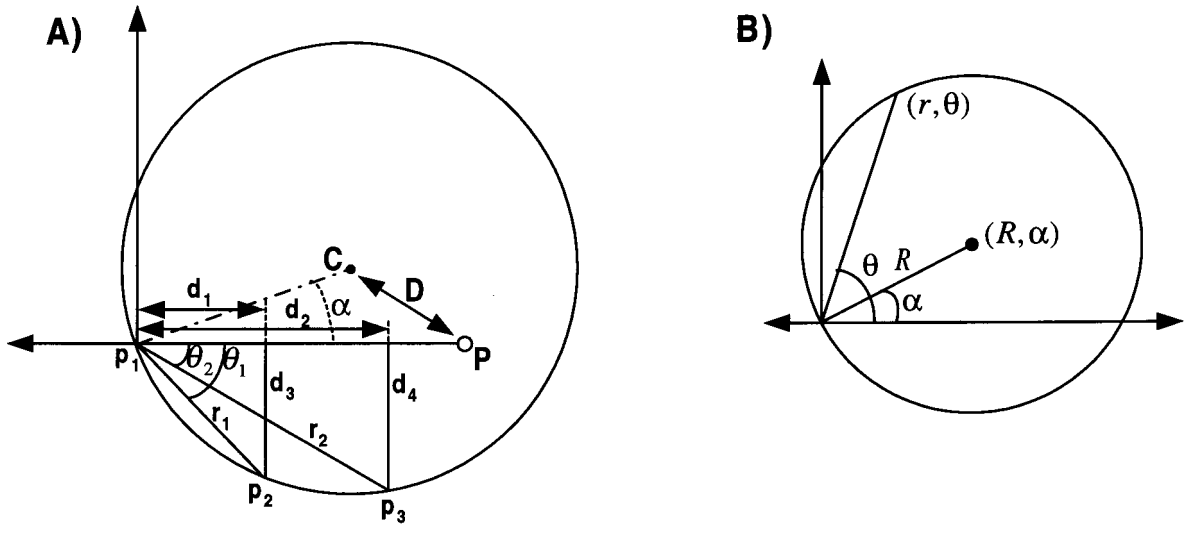

Figure B.1 Determining of centroid C via the perimeter point algorithm.

From A in figure B.1:

$$r_1 = \sqrt{d_1^{\,2} + d_3^{\,2}} \qquad r_2 = \sqrt{d_2^{\,2} + d_4^{\,2}}$$

$$\theta_1 = \operatorname{atan}\left(\frac{d_3}{d_1}\right) \qquad \theta_2 = \operatorname{atan}\left(\frac{d_4}{d_2}\right)$$

From B in figure B.1:

$$r = 2R\cos(\theta - \alpha) \tag{1}$$

$$\cos(\theta - \alpha) = \cos\theta\cos\alpha + \sin\theta\sin\alpha \tag{2}$$

Using (1):

$$r_1 = 2R\cos(\theta_1 - \alpha) \quad (3)$$

$$r_2 = 2R\cos(\theta_2 - \alpha) \quad (4)$$

In order to determine the circle centre, we need R and α:

From (3): $$R = \frac{r_1}{2\cos(\theta_1 - \alpha)} \quad (5)$$

Using (5) in (4): $$r_1\cos(\theta_2 - \alpha) = r_2\cos(\theta_1 - \alpha) \quad (6)$$

Using (2) in (6): $$k_1\cos\alpha + k_2\sin\alpha = 0$$

where $$k_1 = (r_1\cos\theta_2 - r_2\cos\theta_1)$$
$$k_2 = (r_1\sin\theta_2 - r_2\sin\theta_1)$$

$$\therefore \tan\alpha = \frac{-k_1}{k_2} \quad (7)$$

Through (5) and (7) we have the centre coordinate C.

Finally D can be calculated:

$$\overline{D} = \overline{C} - \overline{P} \quad where\ C \Rightarrow (R, \alpha)$$

$$\therefore D = \sqrt{(P_x - (p_{1_x} + R\cos\alpha))^2 + (R\sin\alpha)^2}$$