Implementation of Robotic Visual Attention motivated by Human Physiology and Behavior

Konstantinos P. Michmizos

Department of Biomedical Engineering McGill University, Montreal June, 2006

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Engineering

© Konstantinos Michmizos, MMVI



Library and Archives Canada

Published Heritage Branch

395 Wellington Street Ottawa ON K1A 0N4 Canada Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington Ottawa ON K1A 0N4 Canada

> Your file Votre référence ISBN: 978-0-494-28610-4 Our file Notre référence ISBN: 978-0-494-28610-4

NOTICE:

The author has granted a nonexclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or noncommercial purposes, in microform, paper, electronic and/or any other formats.

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

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

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

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.



Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

To the sacred shadow of my mother, Stamatia and to my father, Polydoros.

К.М.

Abstract

It is Zeus' anathema on biomedical engineering that we should agonise between the Scylla of the simulation of the physiological and anatomical findings and the Charybdis of the efficient implementation of a system. Previous simulations on the addition of the perceptual velocity tracking of a target to a biomimetic controller developed by Galiana allow the implementation of a novel oculomotor system (OCS) in a hardware device comprising two cameras and a neck, each of which able to produce yaw and pitch movement. The robotic system uses visual inputs and a variety of movements in order to track targets, namely slow pursuit (slow phase), saccades (fast phase), conjugate and vergence. The addition of prognosis and memory to the controller, leads to the need for new tactics to be presented that address the problem of fixating a target more efficiently. Since the length and the flatness of the pursuit bandwidth of the new controller are essential for the minimization of the need for saccades, a greater harmonization among the parameters of the controller is basic. The resulting OCS can now be used to label with a significance factor each target presented on the visual field and track the one which is the most conspicuous. The intensity and the speed of the target are the criteria used to select a target and to alternate the alertness state of the robot.

Résumé

C'est l'anathème de Zeus sur la technologie biomédicale que nous devrions agoniser entre le Scylla de la simulation des résultats physiologique et anatomiques et le Charybdes de l'exécution efficace d'un système. Des simulations précédentes sur l'addition de la vitesse perceptuelle de la poursuite d'une cible à un contrôleur biomimétique développé par Galiana a permis l'exécution d'un nouveau système oculomoteur (OCS) dans un dispositif matériel comportant deux cameras et un cou chacun capable de produire des mouvements de rotation autour de l'axe transversal (Yaw) et normal (Pitch). Le système robotique emploie des entrées visuelles et une variété de mouvements afin de dépister des cibles, à savoir poursuite lente (phase lente), saccades (phase rapide), conjugué et vergence. L'addition du pronostic et de mémoire au contrôleur, crée un besoin d'une nouvelle tactique pour adresser le problème de fixer une cible plus efficacement. Puisque la longueur et la planéité de la largeur de bande de la poursuite du nouveau contrôleur sont essentielles pour la minimisation du besoin des saccades, une meilleure harmonisation des paramètres du contrôleur est de base. Le OCS résultant peut maintenant être employé pour marquer avec un facteur de signification chaque cible a présenté sur le champ visuel et suivre la plupart remarquable. L'intensité et la vitesse de la cible sont utilisés comme critères pour choisir une cible et alterner l'état de vigilance du robot.

Acknowledgements

It is a common thing for graduate students to start the written part of their work by thanking their supervisor. This thesis will be no exception. However, I emphasize my eulogy to my supervisor, Dr. Henrietta L. Galiana, not as a consequence of duty but as an indication of my true appreciation to her as an engineer and, more importantly, as a human. Her encouraging guidance and thorough scientific understanding and knowledge of the problems that we had to deal with, illumed her true worry for the outcome of my studies. It is no word of exaggeration to say that this thesis would not be completed without her continuous support.

I feel it as a bounden duty to thank my lab-colleagues as well as Dr. Wagner and Dr. Katsarkas for their support. I should also thank the National Science and Engineering Research Council (NSERC) of Canada, the Fonds Québécois de la Recherche sur la nature et les Technologies (FQRST) and the Onassis Foundation for their financial support of my studies.

A special part of my acknowledgements goes to a good friend of mine, Nicolas Ellaham. For all the discussions in the MNI restaurant, for the numerous hours we spent together discussing from engineering problems to the two natures of God. His wise words will be well remembered.

A special paragraph is kept for the members of my family. I should thank my father, Polydoros, who made numerous overseas trips just to support and see me, my sister, Despina and my brother, Dimitrios.

Finally, I would like to thank all the people who taught me through all these years of my life; inside and outside of the classrooms. I hope one day I will prove worthy of their efforts and worries.

K.M. Montreal, June 2006.

Table of Contents

Chapter 1 Introduction	1
1.1 Motivation and Objective	1
1.2 Thesis Outline	3
Chapter 2 Literature Review -Active Vision	4
2.1 Elements of Visual Perception	4
2.2 Types of Eye Movements	8
2.3 Ocular Reflexes	10
2.3.1 Vestibulo-Ocular Reflex (VOR)	10
2.3.2 Opto-kinetic Reflex (OKR)	12
2.5 Selective Visual Attention in Human	12
2.5.1 The Thalamus	15
2.5.2 The Visual Cortex	16
2.6 Models of the Human Visual Attention System	18
2.6.1 Posner and Petersen's Hypothesis	18
2.6.2 Caltech's Hypothesis	19
2.6.2.1 Tsotsos's extension on the WTA mechanism of the Caltech's Hypothesis.	21
2.6.2.2 Milanese's Extension on the saliency map of Caltech's Hypothesis	22
2.6.3 Other Models	23
2.6.4 Summary	23
Chapter 3 Eye - Head Controller and Robotic Implementation	24
3.1 The bilateral Controller	24
3.1.1 Different Vision Controllers and Robotic Implementations	24
3.1.2 Introduction to the Basic Ideas of the Controller	26
3.1.3 Coordinate System	27
3.1.4 Bilateral Controller	28
3.1.5 Smooth Pursuit Mode	31
3.1.6 Saccade Mode	33
3.1.7 Combining Fast and Slow Phases together	34
3.1.8 Combining Eye and Head Positions: Gaze	36
3.1.8 Combining Eye and Head Positions: Gaze 3.1.9 Vertical Controller	36 37
3.1.8 Combining Eye and Head Positions: Gaze 3.1.9 Vertical Controller 3.1.10 Combination of Horizontal and Vertical Controllers	36 37 38
 3.1.8 Combining Eye and Head Positions: Gaze	36 37 38 40
 3.1.8 Combining Eye and Head Positions: Gaze	36 37 38 40
 3.1.8 Combining Eye and Head Positions: Gaze	36 37 38 40 44
 3.1.8 Combining Eye and Head Positions: Gaze	36 37 38 40 44
 3.1.8 Combining Eye and Head Positions: Gaze	36 37 38 40 44 44 45
 3.1.8 Combining Eye and Head Positions: Gaze	36 37 38 40 44 44 45 46
 3.1.8 Combining Eye and Head Positions: Gaze	36 37 38 40 44 44 45 46 49
 3.1.8 Combining Eye and Head Positions: Gaze	36 37 38 40 44 44 45 46 49 50
 3.1.8 Combining Eye and Head Positions: Gaze	36 37 38 40 44 44 45 46 49 50
 3.1.8 Combining Eye and Head Positions: Gaze	36 37 38 40 44 44 45 46 49 50 e 52

4.4.2 Saccadic Prediction	54
4.4.3 Revised Switching Strategy	55
4.4.4 Reducing Switching Noise	55
4.4.4.1 Velocity Threshold to limit Saccade Duration	56
4.4.4.2 Nullifying Retinal Slip ($R_{\rm S}$) at the onset of the smooth pursuit segment	56
4.4.4.3 Saturation of the Retinal Slip during discontinuities	57
4.5 Final Controller Parameters and Setup Steps for Data Acquisition	57
4.6 Experimental Results	61
4.6.1 Slow Phase	61
4.6.1.1. Controller's Experimental Bandwidth - Head Fixed	62
4.6.1.2. Step Response	02
4.6.1.3. Phase Lag (0.2 Hz and 1 Hz) and Retinal Errors	07
4.6.1.4. Two – dimensional Target	05
4.6.2 Switching Enabled	70
4.6.2.1 Ramp Trajectory	70
4.6.2.2 Two – Dimension Examples	72
4.6.3 Gaze	72
Chapter 5 Towards Target Selection and Visual Attention	79
5.1 Introduction	
5.2 Target Selection	
5.3 Alertness of the Robot	83
5.4 Experimental Results	
5.4.1 Two Targets with Different Velocities and the same Brightness	00
5.4.3 Controller Response under the presence of two targets with different	00
intensities and velocities	89
5.5 Conclusion	91
Chapter 6 Conclusions	92
6.1 Synopsis and Discussion	92
6.2 Significance	93
6.3 Future Extensions	95
6.3.1 Object Recognition	95
6.3.2 Higher Operating Frequency	96
6.3.3 Introducing 3D- Torsion Movement used for Targets with Dimensions	97
6.3.4 Hand – Eye Coordination	97
Appendix A Nonlinear Calibration of the Laser Dot and the Cameras	99
Appendix B Finding Optimal Parameters for PDI Controller	107

List of Figures

Figure 2-1 : Simplified horizontal cross-section of the human eye	4
Figure 2-2 : Fire Rate of an ON Center / OFF surround cell	6
Figure 2-3 : OFF center/On surround and On center/OFF surround ganglion cells	7
Figure 2-4 : Response Characteristic of Neurons to Orientation in the Primary Visual	
Cortex	8
Figure 2-5 : Illustration of typical eye movements.	9
Figure 2-6 : Structure of the human visual system	. 14
Figure 2-7 : Parallel Visual Pathways	. 17
Figure 2-8 : Caltech's Hypothesis.	. 19
Figure 2-9 : Saliency-based model of visual attention	. 21
Figure 3-1 : Bilateral controller conceptual diagram.	. 26
Figure 3-2 : Eye coordinate system	. 28
Figure 3-3 : The symmetric controller.	. 30
Figure 3-4 : Smooth Pursuit Bilateral Controller	. 32
Figure 3-5 : Saccadic Controller.	. 33
Figure 3-6 : The 1-D controller that combines fast and slow phases	. 35
Figure 3-7 : A 1-D gaze model.	. 37
Figure 3-8 : Ocular and Gaze Vertical Coordinate System	. 38
Figure 3-9 : 2-D controller.	. 39
Figure 3-10 : Block diagram of the robotic implementation.	. 42
Figure 4-1 : 1-D PDI Conjugate Pursuit Controller	. 51
Figure 4-2 : 1-D PDI Vergence Pursuit Controller.	. 51
Figure 4-3 : Horizontal Response of the P and the PDI controllers to the chirp input signal	61
Figure 4.4 · Vertical Response of the P and the PDI controllers to the chirp input signs	- UI 1
right 4-4. Vertical Response of the F and the FDF controllers to the entry input signa	. 62
Figure 4-5 : Bode plot of the P and the PDI controllers for horizontal conjugate	
Figure 4.6 . Dodo alot of the D on d the DDI controlling from the line in the	. 63
rigure 4-6: Bode plot of the P and the PDI controllers for vertical conjugate movement	nt.
$\mathbf{F}_{\text{man}}^{i} \wedge \mathbf{A} = \mathbf{C}_{\text{man}} + \mathbf{C}_{\text{man}} + \mathbf{D}_{\text{man}} + \mathbf{D}_{\text{man}} + \mathbf{C}_{\text{man}} + \mathbf$. 64
Figure 4-7: Step Response of P and PDI Pursuit Controller	.65
Figure 4-8: Pursuit Trajectory of P and PDI Controller following a slow target	66
Figure 4-7. Fulsuit Trajectory of P and PDI Controller following a rapid target	00
Figure 4-10.2 D target D controller number to 0.2 HZ input	08
Figure 4-11. 2 D target P controller pursuit vs. time for 0.2 Hz input	68
Figure 4-12:2 D target pursuit of PDI controller to 0.2 Hz input.	69
Figure 4-15. 2 D target PDI controller pursuit vs. time for 0.2 Hz input.	69
Figure 4-14 : Response of the P and PDI controller to a horizontal conjugate ramp input	ut
with low speed.	71

Figure 4-15 : Up: The PDI controller's conjugate eye response to a fast moving ramp. Down: The complete experiment showing the sinusoidal target trajectory and the	
conjugate response to it	2
Figure 4-16 : 2 D target of P controller to 0.2 Hz input with saccades allowed	3
Saccades enabled. $/$	5
Figure 4-18: 2 D target of PDI controller to 0.2 Hz input with saccades allowed	4 1
saccades enabled	4
Figure 4-20 : Horizontal Gaze Response of the P controller	6
Figure 4-21 : Horizontal Gaze Response of the PDI controller	6
Figure 4-22 : Horizontal Gaze Response of the PDI controller to a fast moving sinusoid	
target	7
Figure 5-1 : The Bottom-Up control of Visual Attention Model used for target selection.	
8	1
Figure 5-2 : Flow- Chart of the Target-Selection Mechanism 8	2
Figure 5-3 · Image Pathway	Δ
Figure 5-4 : Manning of Mean Pixel Blob Intensity to K and K	- -
Figure 5-5: 3 D plot of the experimental response of the horizontal surguit controller	5
under the presence of two terests	7
under the presence of two targets	1
Figure 5-6: Individual Horizontal and Vertical Dimension of the experiment shown in	_
Figure 5-5	8
Figure 5-7 : 2-D plot of the experiment shown in Figure 5-5	8
Figure 5-8: Step Response of the Pursuit Horizontal Controller for Various Intensities	
Step Inputs	9
Figure 5-9 : Experimental Response of the Horizontal Pursuit Controller under the presence of two targets with different intensities and velocities	0
Figure A-1 : Laser Beam Scattering	0
projected to the wall 10	1
Figure A-3 : Mapping Functions for Horizontal (left column) and Vertical (right column)
Pixel distance (first row) and laser dot position (second and third row)	4
Figure A-4 : Simulated Sinusoid Input Voltage received from the Function Generator	•
(up) and the output of the mapping function (down) in angles 10	5
Figure A-5 · Sample Error in Pixels (top papel) and in degrees (bottom level) 10.	6
Figure <i>I</i> is a sumple birde in trivels (top patier) and in degrees (bottom lever)	0
Figure B-1 : Horizontal Vergence PDI Controller Bondwidth (unner nanch) Stan	
Response (left bottom panel) and Deak to Deak abachite avay (under) sheet (-i-t-t-t-t-t-	
need with respect to the propertional gain (D)	0
Figure D : Horizontal Vargance DDI Controller Devil 1111 (5
Figure D-2 : Horizonial vergence PDI Controller Bandwidth (upper panel), Step	
Response (left bottom panel) and Peak-to-Peak absolute over(under)shoot (right bottom	
panel) with respect to the slip gain (D)	9

Figure B-3 : Horizontal Vergence PDI Controller Bandwidth (upper panel), Step
Response (left bottom panel) and Peak-to-Peak absolute over(under)shoot (right bottom
panel) with respect to the integral gain (I) 110
Figure B-4 : Horizontal Vergence Controller Bode plot for old parameter set (red dashed
line) and new parameter set (blue solid line)
Figure B-5 : Horizontal Vergence Controller Step Response for old parameter set (red
dashed line) and new parameter set (blue solid line)
Figure B-6 : Horizontal Vergence Controller Percentage of Peak-to-Peak
over(under)shoot for old parameter set (red dashed line) and new parameter set (blue
solid line)
Figure B-7 : Horizontal Conjugate Controller Bode plot for old parameter set (red
dashed line) and new parameter set (blue solid line)
Figure B-8 : Horizontal Conjugate Controller Step Response for old parameter set (red
dashed line) and new parameter set (blue solid line)
Figure B-9 : Horizontal Conjugate Controller Percentage of Peak-to-Peak
over(under)shoot for old parameter set (red dashed line) and new parameter set (blue
solid line) 113

List of Tables

Table 4-1: Table of parameters used for the horizontal OCS controller in the P controller
58 Table 4-2 : Table of parameters used for the vertical OCS controller in the P controller 58
Table 4-3 : Table of parameters used for the horizontal OCS controller in the PDI controller 59
Table 4-4 : Table of parameters used for the vertical OCS controller in the PDI controller
Table A-1 : Horizontal Dimension Calibration Measurements 102
Table A-2 : Vertical Dimension Calibration Measurements
Table A-3 : Parameters of the Mapping 3 rd degree Polynomial Functions
Table B-1 : New Parameter Set for Conjugate and Vergence Horizontal and Vertical Controller 110

Chapter 1 Introduction

1.1 Motivation and Objective

As we drive from work to home, we pay attention to the other drivers' behavior, we change lanes and we keep an eye on the speedometer while at the same time tuning the radio to our favorite song. At home, we park the car in a narrow space and as we walk home, in the dark, we are able to recognize our neighbor by the way he walks. How do we cope with these kinds of situations? And the most challenging question to answer would be, how could a robot deal with them?

Recent advances in humanoid robotics, such as the introduction of Honda's Asimo robot [118], have brought humanoids to the forefront of robotics research. In the future, a biomimetic robot may serve as another set of eyes, ears, hands and legs for all kinds of people in need. Someday it might help with important tasks like assisting the elderly or a person confined to a bed or a wheelchair. It might also perform certain tasks that are dangerous to humans, such as robot assisted surgical operations, space explorations, fighting fires or cleaning up toxic spills.

Though today's computers can perform billions of operations per second, they are still no match for even a young child when it comes to skills such as pattern recognition or visual processing [90]. Robotic vision introduces requirements for real-time processing and removes convenient assumptions in static camera systems that allow simplifying computations. An area in which simplification and computational speed are crucial is that of the control of the eyes – head robotic systems. The biomimetic bilateral controller, first developed by Galiana [35], is based on the physiological evidence that smooth pursuit - that is the slow movement of the eyes - and saccade – their rapid movement - are handled by the same neural circuitry. Therefore, it is an ideal candidate

to be placed in the heart of a humanoid robotic system, to illustrate principles of parsimony and simplicity.

The first implementation of the bilateral controller was conducted by Wagner [103] and consisted of a binocular robot that was run on a UNIX system. Both horizontal and vertical motions were available, the head was not moving but the controller accommodated both pursuit as well as saccades. The eyes where modeled by analog position sensing detectors (PSD) actuated by PID controlled motors. No image processing was available, constraining the target to being a very bright laser dot.

The first implementation of the bilateral controller using digital cameras to model the eyes was conducted by Palatsoukas [73]; it also introduced the use of the head to model head-free gaze control. The image processing was simplified since it was guaranteed that the target would only run along a thin strip in the horizontal direction.

The second implementation of a binocular system with horizontal and vertical movements using digital cameras was done by Sejean [91]. The two eyes were mounted on a neck platform that provided the yaw and pitch required for 2D head-free gaze. For the first time, the head was equipped with an angular velocity sensor capable of providing gaze stabilization for head movements. The entire image obtained from the digital camera was analyzed to take the center of gravity of pixel intensities.

Lee [60] improved upon the model in Séjean's work, which only used position error signals, by following biological cues and incorporating derivative and integral signals in the slow-phase pursuit control. At the same time, derivative signals were also added to fast-phase control, as a form of primitive prediction. This study was restricted to theoretical simulations, based on pre-identified hardware models.

Lee's work was purely theoretical and its value was tested only in simulation. This thesis will first deal with the implementation of that work in actual hardware and will describe the creation of a binocular robotic system with horizontal and vertical movements embodying derivative and integral signals in the pursuit and the saccadic modes. The core work of this thesis concerns the addition of image processing capabilities in a robot operating in an unknown environment and using point features in the world as visual landmarks. This addition allows for the first time the selection of the target to be made from the superimposition of the intensity image and its derivative image weighted accordingly and the experimentation with two targets within the field of view of the robot.

1.2 Thesis Outline

The journey through this thesis starts in *Chapter 2* in which a basic understanding of human visual perception is developed. After a short presentation of the important structures involved in ocular tracking, the properties of the visual system that can guide attention are discussed. Various models are analyzed in detail and the relationship between them and the neurobiology of primate visual attention mechanism is presented.

Chapter 3 describes the theory behind the bilateral controller and its core butterfly concept. It also introduces the mechanisms of the eye-head gaze system and concludes with a presentation of the various components used for the experimental implementation.

Chapter 4 analyzes modifications made to the slow phase controller as well as to the fast phase controller and shows the results of the modified controller implemented on the robot. Switching noise between modes is also addressed in this chapter.

Chapter 5 introduces the perceived intensity and speed of the target to attribute a significance factor to it that allows its selection in a visual scene containing multiple distractors. The intensity of the target is also embodied into the gains of the controller altering the alertness of the robot with respect to the brightness of the target.

Chapter 6 discusses the significance of the presented results. The thesis concludes with a proposed series of possible future modifications and extensions to the system.

Chapter 2 Literature Review -Active Vision

This chapter will summarize the elements of biological visual processing and the resulting eye movements that can be observed. Models for different ocular reflexes will also be described.

2.1 Elements of Visual Perception

The eye is a near sphere, with an average diameter of approximately 20 mm. The membranes that enclose the eye are: the *cornea* and *sclera* outer cover, the *choroid* and the *retina*. The choroid coat is heavily pigmented and thus helps to reduce the amount of extraneous light entering the eye as well as the backscatter within the optical globe. At its anterior extreme, the choroid is separated into the *ciliary zonule* and the *iris* diaphragm. The iris contracts or expands in order to control the amount of light that enters the eye. The lens is made up of concentric layers of fibrous cells and is suspended by fibers that attach to the ciliary zonule. The lens absorbs approximately 8% of the visible light spectrum, with relatively higher absorption at shorter wavelengths.



Figure 2-1 : Simplified horizontal cross-section of the human eye. (Adapted from [113])

The innermost membrane of the eye is the retina, which lies on the inside of the wall's entire posterior portion. When the eye is properly focused, light from an object outside the eye is imaged on the retina. *Pattern vision* is afforded by the distribution of discrete light receptors over the surface of the retina. There are two classes of receptors: *cones* and *rods*.

The cones are located primarily in the central portion of the retina, called the *fovea*, and are highly sensitive to color. Humans can resolve fine details with these cones mainly because each one is connected to its own nerve end. Muscles controlling the eye rotate the eyeball until the image of an object of interest falls on the fovea. The rods are distributed over the retinal surface. The larger area of distribution and the fact that several rods are connected to a single nerve end reduce the amount of detail discernible by these receptors.

By taking some liberty in interpretation, we can view the fovea as a square sensor array of size 1.5 mm x 1.5 mm. The density of the cones in that area of the retina is approximately 150,000 elements per mm^2 which is certainly within the realm of current electronic imaging sensors.

The retina consists of seven layers of cells. The last layer is the ganglion cell layer which has a special role. Their main cell body resides in the retina but they also have an efferent axon and they use that long "cable" to transmit pulses to various areas in the brain. There are about one million ganglion cells, and each has one *axon*. The axons from each eye bundle together to form the *optic nerve*. The optic nerve carries the encoded retinal image to the brain. Each ganglion cell sends a sequence of pulses ("spikes") from the retina to the brain. The response of the cell is often summarized by the pulse rate. For any retinal ganglion cell there is a small region of the retina that affects the response rate of that cell which is called the *receptive field* of the cell.

Kuffler ([55], [56]) found ganglion cells for which the response increased when a tiny spot of light appears in a particular region. This was called the *excitatory region*



Figure 2-2: Fire Rate of an ON Center / OFF surround cell

for that cell. He also found that in the periphery of that area there was an annulus-shaped region in which the tiny spot of light decreased the pulse rate of the cell. This surrounding region is called the *inhibitory region*. Because these cells were excited by light in the center and inhibited by light in the annulus surrounding the center, these cells are called *On center / OFF surround*.

Kuffler also found retinal ganglion cells that had the opposite property, namely there was a central round region in which the electrical activity of the cell was inhibited by a visual stimulus, and a surrounding annulus region in which the cell was excited. He called these cells *OFF center / ON surround*.

Moving further into the brain, signals from the retina reach an area near the center of the brain called the *lateral geniculate nucleus* (LGN). Cells in the LGN then relay these signals to the back of the brain (back of the head). The surface of the brain in general is called the *cortex* and the surface of the brain at the back of the head is called *primary visual cortex* (V1) because this is the first area of the cortex to receive visual inputs.



Figure 2-3: OFF center/On surround and On center/OFF surround ganglion cells

The LGN cells do more than simply retransmit the same signals. In particular, although the LGN receives axons from the retinal ganglion cells (about 10^6), it receives far more axonal inputs (about 10^7) from the visual cortex. Yet, it is not understood how this feedback loop works [69].

The first experiments to examine cells in the primary visual cortex were carried out in the late 1950's [43]. The responses of single cells in the cortex of anaesthetized cats were examined and what was found is that each cell responds to a small area of the visual field. Unlike ganglion cells in the retina, the responses are not radially symmetric. Instead the cells are tuned to a particular orientation. The response of this kind of cells, called *simple cells*, can be thought of as a weighted average of the image intensity over an ellipsoidal region.

The weights are positive along a center stripe and negative along the two flanking stripes. These cells can be thought of as *line detectors* and are involved in detecting lines in images, more generally, detecting oriented structure.

Other cells that are tuned for motion and their role will be linked to the slip signals used in the controller and will be presented in Section 4.2.



Complex Cell is excited by all three stimuli

Figure 2-4 : Response Characteristic of Neurons to Orientation in the Primary Visual Cortex

2.2 Types of Eye Movements

A hundred years ago, it was already established that "there seems to be an almost ceaseless twitching, as if rest for more than an instant were the one thing not to be endured" [94]. Eye Movements are arguably the most frequent of all human movements. This phrenetic movement is a consequence of the colossal amount of visual information that is available. Instead of devoting resources to processing it all in detail, evolution appears to have selected a solution whereby small portions of the visual world are inspected in a rapid sequence.

The entire goal of the Oculomotor System is to keep the object of interest on the fovea [21]. Two general classes of eye movements have been distinguished: *Conjugate* eye movements and *vergence* eye movements. Conjugate eye movements are those that preserve the angular relationship between the right and left eyes. Up and down movements and combinations of vertical and lateral movements also fall into the conjugate category. Vergence eye movements are ones where the angle between the eyes changes. This is necessary for depth perception as well as focusing on objects at various distances. It also permits both eyes to simultaneously center the target on their respective retinae.



Figure 2-5 Illustration of typical eye movements. Eyes move in opposite direction for vergence, same direction for conjugate commands. Saccades rapidly drive both eyes to new location.

In the context of target tracking, conjugate eye movements are of two types: *smooth pursuit* and *saccades*. Pursuit permits the visual system to keep a moving object on the fovea. This is necessary because the visual target should be viewed with the most precision. To do this, it is believed that pursuit uses the retinal error (the distance between the target and the fovea), and slip (the derivative of retinal error), as the main error signals in negative feedback loops [83]. Pursuit is maintained as long as the image moves reasonably slowly: if the retinal error increases too quickly, a saccade is executed to quickly reduce the retinal error so that the objects may be analyzed in detail.

Saccadic eye movements are used to rapidly orient the eyes towards the target. Saccades are executed blindly: vision is not used in a negative feedback loop as in pursuit ([78], [79]). These rapid movements exploit the foveated structure of the eye, quickly reorienting it to place a target near the sensitive fovea. Depending on target trajectories, multiple sequences of slow pursuit with saccades are observed, generating *pursuit* *nystagmus*. Figure 2-5 illustrates how the eyes move in response to conjugate, vergence, and saccadic commands.

All binocular eye movements can be divided into both a conjugate and a vergence component. The bilateral controller that will be presented in depth in Chapter 3 exploits this fact. This means that no matter what the speed of a given eye movement, left and right eye position can be mapped from their conjugate and vergence contributions.

These are intentional eye movements generated in response to the appearance of targets of interest. However it is possible to generate such eye movements reflexively (unconsciously) when exposed to certain types of head or visual field perturbations.

2.3 Ocular Reflexes

2.3.1 Vestibulo-Ocular Reflex (VOR)

The vestibulo-ocular reflex (VOR) is a primitive eye-movement reflex driven by head movement that stabilizes images on the retina. The greatest potential source of image slip on the retina is due to self-rotation [76]. By producing an eye movement in the direction opposite to the head movement, the VOR stabilizes automatically vision relative to fixed space. Since slight head movements are present all the time, the VOR is very important for stabilizing the image on the center of the visual field. Patients whose VOR is impaired are not able to read, because they can not stabilize the eyes during small tremors. The VOR reflex does not depend on visual inputs and works even in total darkness or when the eyes are kept closed.

The sensory drive in the VOR is the vestibular system: The dynamic changes in linear acceleration of the head are sensed by the otoliths, composed of the saccule and the utricule and by the semicircular canals. The otoliths are located inside the semicircular canals and provide 3D sensing of *linear* acceleration. The semicircular canals are nearly orthogonal sensors located in the inner ear and react to *angular* rotation in the three dimensions, acting like biological accelerometers. The ampulla is the enlarged portion of

each semicircular canal that possesses the sensory areas of this vestibular system. The cupula which is a gelatinous membrane located within the ampulla of each semicircular canal, contains the cilia of the semicircular canal hair cells and displaces as it experiences angular acceleration.

Hair cell movement is the basis of both types of sensors: the sensor geometry then selects angular or linear motion [106]. Perception is achieved through mechano-neural transduction: When the head begins to rotate, experiencing angular acceleration within the plane of a semicircular canal, the inertia of a viscous fluid filling the membranous labyrinth called endolymph causes it to remain stationary while the canal rotates with the head. This causes fluid to flow from the duct into the ampulla, forcing the cupula to bow like a drumhead in the direction opposite to that of the rotation, thus, deflecting the cilia of the hair cells. As the head continues to rotate at a steady velocity, viscous forces between the canal and the endolymph cause the endolymph to "catch up" with the canal, eliminating the relative movement. Elastic properties of the cupula then return it to a vertical position. With the cessation of spinning (angular deceleration), the moving fluid pushes against a suddenly still cupula, and the cupula is deflected in the opposite direction. It is in this manner that a semicircular canal senses rotational acceleration of the head about its particular axis of rotation. However, if the canal is moved along a linear axis, experiencing linear acceleration, the endolymph will move with the canal because it is incompressible. In this case the cupula will not bend. To summarize, the polarization of the hair cells leads to an altered frequency of action potential discharge of cranial nerve VIII which is the efferent pathway for vestibular signals, transmitting head movement and head positioning data to various centers in the brain and postural control nuclei.

The VOR is mediated by a trineuronal arc composed of the primary vestibular neurons, relay neurons in vestibular nuclei in the brainstem, and oculomotor nuclei, which contain motoneurons that drive eye muscle activity. Axons entering the medial vestibular nucleus synapse with neurons that send their axons to the contralateral abducens nucleus. This nucleus consists of neurons whose axons comprise cranial nerve VI and innervate the lateral rectus muscle. Other cranial nerve VIII axons enter the lateral vestibular nucleus. These neurons relay to the ipsilateral ocular motor nucleus whose neurons form the axons of cranial nerve III which innervates the medial rectus nucleus muscle. The cerebellum is essential for motor learning to correct the VOR in order to insure accurate eye movements.

The vestibular information that is transmitted to the brain through the vestibular nerve is encoded into modulations in neural firing rate. One or both eyes may exhibit an involuntary, rhythmical, repeated oscillatory pattern in any or all fields of gaze, called *nystagmus*. If head rotation occurs in a completely dark environment, the nystagmus is due purely to vestibular input and is called *vestibular nystagmus*.

2.3.2 Opto-kinetic Reflex (OKR)

The optokinetic eye-movement response (OKR) is another ocular reflex induced by visual stimuli using movements of large visual fields relative to the head. In order to compensate for the motion of visual fields, the optokinetic system uses the relative velocity of the image on the retina to induce eye movements in the same direction and at the same velocity as the external world to preserve stable vision.

VOR and OKR share the same vestibular nuclear neurons and oculomotor neurons, and synergistically stabilize vision. The VOR/OKR pathways rely on extensive commissural connections between the bilateral vestibular nuclei, internuclear connections between motor nuclei, and other connections with the tegmentum and the cerebellum ([81], [36]).

2.5 Selective Visual Attention in Human

As you read this sentence, your visual system must not only identify each word, it must simultaneously ignore myriad competing visual stimuli, from the words on the rest of this page to the leaves rustled in the wind outside your window. The selection of a subset of sensory signals for preferential processing is known as *attention*. Attentional filtering of sensory input is necessary because our sensory systems are continually inundated with information from different stimuli, any of which can potentially be used to guide behavioral responses. The brain must therefore concentrate its limited resources on analyzing the most important aspects of the sensory scene.

Human vision can basically be divided into two main phases, *low-level vision* and *high-level vision* [64]. Although the frontier between the two vision phases is not clearly stated, their main roles have already been established.

Low-level vision starts with gathering visual information by the means of more than 200 million retina photoreceptors on a rate of 3 bit/s for each neuron [97]. The gathered information is then transmitted to the visual cortex where depth, color, motion, orientation, and shape information is extracted. Then the high-level vision performs its task on these extracted features. It is mainly responsible for recognizing scene contents by matching the representative scene features to a colossal database of learned and memorized objects (object recognition). Despite the huge amount of visual information to be processed and despite the combinatorial nature of the recognition task, it has been estimated that humans can recognize a large number of objects in less than 200 ms [97].

The existence of a mechanism that selects only a reduced set of the available visual information for high level processing seems to be the most coherent explanation of the high performance of the human visual system [101]. This mechanism is referred to as *visual attention*, the ability of vision systems to rapidly select the most salient and thus the most relevant data in a scene, leading, thereby, to a considerable speed up of the entire vision process. The visual attention mechanism seems to control the selection of the informative parts of the scene to which the fovea is oriented.

Through the optic nerve, the visual information collected by the retina is transmitted to two different brain centers; namely the *lateral geniculate nucleus* (LGN), which is a part of the thalamus, and the *superior colliculus* (SC). The visual information flows into the various brain areas using two main pathways, namely the *retino-geniculate* and the *collicular*. About 90% of the visual information is transmitted by the retino-geniculate pathway into the visual cortex. This information includes spatial, temporal,



Figure 2-6 Structure of the human visual system. The two major information flows from the retina to the visual cortex (modified from [116])

chromatic, and disparity processing. Since the collicular pathway is only responsible for the remaining 10% of the visual information, it is often considered negligible in terms of describing in general terms the human visual system. However, because of its important role in visual attention and eye movements [31], it will be analyzed further.

The retinal photoreceptors are connected with the ganglion cells via bipolar cells. The ganglion cells can be classifed into three general categories: the M (Magno) ganglion cells that have large dendritic arrays, and receive information from a wide radius of bipolar cells. They are mostly found in the peripheral retina where they connect with cones, are not color-sensitive, and are "coarse-grained" and relatively insensitive to detail. Their main asset is that they are sensitive to motion. The P (Parvo) ganglion cells

connect with foveal cones and thus are associated with central vision. They are small ganglion cells that dominate in the fovea: they are color sensitive and are "fine-grained", meaning their receptive fields are small enough that they can pick up a high level of detail. Since they connect with foveal cones, they are associated with central vision. A third category which embodies the *non-M non-P ganglion cells* has also been identified. These cells have properties similar to both the M and P cells.

2.5.1 The Thalamus

Regarded as the "Rosetta Stone" of the nervous system for a long time now, the thalamus is the nucleus of a huge amount of relaying activity, giving a lot of insight into the importance and direction of various neural signals. It is a routing station for all incoming sensory impulses except those of smell, transmitting them to higher (cerebral) neural centers. In addition, it connects various brain centers with others [107].

The thalamus is located in the anterior and dorsal side of the midbrain and is composed of three nuclei: the *lateral geniculate nucleus* (LGN), the *pulvinar nucleus*, and the *reticular nucleus* (RN).

The LGN forms the main relay of visual information to the cerebral cortex. It is a cortical area that segregates the various retinal subsystems serving the contralateral visual fields and organizes their projections to the visual cortex via the retino-geniculate pathway. The cells constituting the LGN are of three different types: *parvocellular*, *magnocellular* and *koniocellular*. The parvocellular and the magnocellular cells are synapsed by the P and M ganglion cells of the retina respectively. These two types of cells feed the parvocellular and the magnocellular visual stream of the retino-geniculate pathway [61]. Recent works [40] proved the existence of a third stream -the koniocellular- related to the non-M non-P ganglion cells of the retina that feed into the koniocellular cells of the LGN.

The pulvinar is the largest nucleus of the thalamus, since it occupies approximately 40% of its entire volume. Although it has no direct input from the optic nerve, the pulvinar has reciprocal connections with all cortical areas. It is believed that this nucleus plays an important role in attention guiding and eye movements [58].

The attention mechanism is also affected by the reticular nucleus (RN). The cells of this nucleus seem to be modulated by an alerting-related signal that is a signal generated during an alert state in order to accelerate the selection of the visual information and thus to allow for instance rapid reaction to danger.

2.5.2 The Visual Cortex

The neurons in the LGN send their axons directly to V1 (primary visual cortex, striate cortex, area 17) via the optic radiations. This highway of visual information courses through the white matter of the temporal and parietal lobes. Once the axons reach V1, they terminate primarily in a single sub-layer of cortex.

The about 10¹⁰ cells of the visual cortex are organized in a hierarchical manner. The area V1 is at the basis of the hierarchy, since it represents the major entry points of the visual information coming from the LGN to the visual cortex. It is noteworthy that half of the area of V1 is devoted to the representation of the visual information gathered by the central part of the fovea. From V1, the visual information is then transmitted to the area V2. There-after, the two visual pathways seem to diverge, since the parvocellular stream feeds to V4, whereas the magnocellular one feeds to Middle Temporal area (MT).

V4 is associated with the synthesis of complex forms and with detection of color and form. It transmits the processed visual information, via the parvocellular stream, to the inferior temporal cortex (IT), which represents the final visual area for object recognition. For this reason, the parvocellular stream is also known as the *what* stream. The middle temporal (MT) area represents an important route of visual information to the posterior parietal cortex (PP), which is responsible for object localization. Thus, the



Figure 2-7 : Parallel Visual Pathways and their suggested functions in the macaque monkey (modified from [48])

magnocellular pathway is often referred to as the *where* stream. The *where* stream is believed to play a fundamental role in controlling visual attention deployment [47].

These are several streams for visual information which can cause relocation of the visual axis either by reflexive (unconscious) mechanisms or by conscious analysis of a visual scene and target selection. In the following section, the "where" pathway is further analyzed in the sense of mathematical models that are based on the physiology of the human brain and use significance labels to distinguish between the distractors and the target.

2.6 Models of the Human Visual Attention System

Imagine you're looking down the concourse in a Central Metro Station and you have been asked to find a lady in a red dress. Of course, there are hundreds of surrounding people moving in different directions. However, there seems to be no problem at all in detecting the woman in the red dress walking along. The visual system uses the dress's color to filter out all the irrelevant noise around it and homes in on the moving object of interest.

Since Hubel and Wiesel's pioneering studies [43] in the visual cortex, most visual scientists have assumed that the perception of form, color, depth, and motion corresponds to the firing of cells specialized to detect these visual qualities.

Recent advances on functional Magnetic Resonance Imaging (fMRI) clarified that eye movements are driven both by properties of the visual world and processes in a person's mind. This led to the establishment of some hypotheses with respect to the organization of the visual attention system. In the following, we briefly describe three of these hypotheses which follow the general argument that the visual attention system is distributed over a network of brain areas. Furthermore, all three hypotheses point out that most of the brain areas which control the visual attention mechanism are also involved in the control of eye movement.

2.6.1 Posner and Petersen's Hypothesis

According to Posner and Petersen [76], visual attention involves three operations in the following sequence: firstly, there is a need of disengagement of the attention from its current locus; secondly the attention must be moved to the new target; and finally it must be engaged on the new selected location. In [76], it is suggested that the posterior parietal cortex (PP) controls the disengagement of the attention, whereas the move step is controlled by the midbrain, in particular by the superior colliculus (SC). The engagement operation depends on the pulvinar nucleus of the thalamus.



Figure 2-8 : Caltech's Hypothesis (modified from [52])

2.6.2 Caltech's Hypothesis

Koch and Ullman [52] originally and Niebur [70] subsequently elaborated one of the most popular hypothesis on visual attention. Caltech's hypothesis –as it was namedrepresents one of the most tangible analyses of the cortical areas entailed in the mechanism of controlling the visual attention. The authors went further to elaborate a computational model that matches the hypothesis, namely the *saliency-based model of visual attention*.

According to this hypothesis, elementary features of the image are extracted in cortical areas within and beyond the striate cortex V1. For example, motion is analyzed by MT and color is extracted by V4. The saliency of locations over the entire visual field is represented in a unique scalar map, called the *saliency map*, which resides either in LGN or in V1 and corresponds to the master location map of the feature integration theory. Furthermore, the saliency of scene locations is strongly influenced by the surrounding context. The *Winner-Take-All* (WTA), which is believed to be located in the thalamic reticular nucleus, and the *inhibition of return*, suggested to be implemented by

multiscale center-surround filters, are the mechanisms that drive the attention over the visual field.

Concerning the inhibition of return, Allman et al. discovered that many MT cells are able to integrate motion information from a large swath of the scene. "Even though an MT cell may respond directly to just one spot in the visual field," says Allman [2], "the cells have knowledge of what's going on in the region surrounding them."

The model commences by extracting, in parallel, a set of scene features like color, orientation, motion etc. In each one of the parallel computational branches, a conspicuity map is created for each considered feature using a lateral inhibition mechanism. Hence, each conspicuity map highlights the parts of the scene that firmly differ, according to the corresponding feature, from their surroundings. The parallel conspicuity maps are then superimposed into a single map of attention, the saliency map, which topographically encodes for location saliency over the entire scene and with respect to all considered features. After the construction of the saliency map, a further step of the model is required in order to find the most salient scene locations: A maximum network, which is generally implemented using a Winner-Take-All (WTA) network, allows the selection of the most salient locations often referred to as *Focus of Attention* (FOA).

The selected region (winner) is considered as the most salient part of the image. The focus of attention is then shifted to this location. Local inhibition is then activated in the saliency map, in an area around the FOA. This yields dynamical shifts of the focus of attention by allowing the next most salient location to subsequently become the winner. Besides, the inhibition mechanism prevents the spot of attention from returning to previously attended locations and hence, creating instability in the model. The number of detected locations can be either set by the user or determined automatically through the activities of the saliency map.



Figure 2-9: Saliency-based model of visual attention that represents feature extraction, conspicuity computation for each feature, saliency map computation by integrating all conspicuity maps and the detection of the focus of attention by means of a Winner – Take – All Mechanism.

2.6.2.1 Tsotsos's extension on the WTA mechanism of the Caltech's Hypothesis

There are a large number of computational models of visual attention that have been inferred from the hypothesis stated by Koch and Ullman. Under the assumption that a saliency map does exist and is already calculated, Tsotsos et al. [102] concentrate on the Winner-Take-All mechanism.

A processing hierarchy is introduced which can be regarded as a pyramidal representation of the already calculated saliency map. At each layer of the pyramid, the units are computed as a weighted sum of certain neighbors from the underlying layer. The hierarchy is composed of computing units (referred as *interpretive units*) which perform processing related directly to the interpretation of their input (e.g., color, edges,

motion). Each interpretive unit receives feedforward as well as feedback connections within the pyramid. The globally winning unit activates another WTA that operates only over its directly connected feedforward inputs. This concentrates the winning unit's location within the top-level winning receptive field. Pruning inhibits all of the feedforward connections of the pyramid that do not contribute to the winner.

At the top level of the pyramid a "beam", with a certain radius, is projected around a winner location (e.g. the location with the highest activity) and expands as it traverses the hierarchy. A WTA mechanism is activated at each level in order to drive the beam into the next level. The global winner is then determined at the lowest level of the hierarchy.

2.6.2.2 Milanese's Extension on the saliency map of Caltech's Hypothesis

Milanese ([66], [67]), based on the work of Posner [76], extended Caltech's hypothesis to a more detailed description of the human visual attention system. He agreed on the fact stated by Koch, that the cortical areas are the ones responsible for computing feature maps (color, orientation, motion), and that there is integration of all these maps into a saliency map. He also reckoned the superior colliculus as the brain area that guides the integration process and the pulvinar and the reticular nuclei of the thalamus as the areas in which the final map resides.

What was also experimentally found is that top-down or task dependent influences on the attention behavior seem to be received by the SC from the inferior temporal cortex (IT). Furthermore, it is reported that the M-cells of the retina directly project alert signals onto the SC and that the parietal cortex (PP), with the ability to estimate spatial relations, can predispose the selection of new targets in reference with their spatial relation with previously attended locations [66].

The model, assuming an already existing ideal WTA mechanism that selects by default the most salient location of the scene, generates a set of feature maps for features like color, intensity and orientation which are extracted from a color input image. At the following step, each feature map is transformed into a conspicuity map- using a bank of Difference of Oriented Gaussian filters (DoOrG) that discriminates outstanding regions according to a specific feature. Finally, the conspicuity maps are integrated into the final saliency map, after undergoing a non-linear relaxation process.

2.6.3 Other Models

It is obvious that a significant number of other works, which have dealt with the computational modeling of visual attention, exists. VISIT [1] is a connectionist model of attention which involves bottom-up as well as top-down mechanisms to select salient objects in a scene. The Guided Search model [108] is another example that integrates image-based stimuli and task-dependent knowledge into an overall activation map that corresponds to the saliency map. Olshausen *et al.* have presented a computational model that simulates the *shifter circuits theory* [71], that is the segregation of objects of interest and the routing of the corresponding visual information to higher stage of the visual cortex. Privitera and Stark [74] have proposed a model of visual attention, whose internal parameters (used features, feature weights etc.) can be adapted to the type of analyzed image. The purpose of the work was to reproduce the human scan path with a computational model. Other computational models include a feature integration theory [98], SERR [44], Dynamic Routing Circuits [72], What –and –Where filter [22], Active Vision [3] etc.

2.6.4 Summary

Various visual attention schemes have been used to model visual attention. A common aspect throughout all the works that have been mentioned before is the superposition of the pre-defined features of the image to form a universal image that designates the selection the target. This idea of weighting and summing up different parameters of a random image will be used in Chapter 5 in which a selection between a target and a distractor will be made possible.

Chapter 3 Eye - Head Controller and Robotic Implementation

3.1 The bilateral Controller

The last chapter introduced most of the various physiological and anatomical findings on which the physical robotic implementation bases its efficient operation as well as the mathematical models that will be used for target selection. This chapter, after presenting the most important previously implemented robotic vision systems, describes the heart of the eye-head controller on which the different control and target selection strategies are tested.

3.1.1 Different Vision Controllers and Robotic Implementations

This section will provide a brief discussion of other robotic vision systems. Most of the projects either focus on exploiting the image processing side, or the results and they do not specify a parameter set in achieving desired responses.

TRICLOPS ([T]he [R]eal-time, [I]ntelligently-[C]ontro[L]led [O]ptical [P]ositioning [S]ystem) is a robotic head designed at the National Institute of Standards and Technology (NIST) ([34], [104]). It is categorized as a kinematic controller: Inverse kinematic control is mostly used for control of robotic limbs.

ESCHeR (E[tl] S[tereo] C[ompact] He[ad for] R[obot Vision]) is a compact and light robotic head, providing six degrees of freedom ([86], [57]). ESCHeR's wide-angle foveated lenses are unique: unlike other systems, it combines the need for a highly sensitive fovea, while still providing peripheral vision with less acuity (up to 120 degrees.)
Developed at the University of Coimbra, MDOF ([M]ulti-[D]egree [O]f [F]reedom) is a binocular vision system designed to replicate human visual movements ([11], [9]). MDOF uses separate controllers for vergence and saccades. There is also a separate controller for pursuit. The controller uses the target coordinates in space to output the appropriate motor drives. Both position and velocity are servoed, or controlled within the feedback loop. Again, inverse kinematics combined with the motor drives are used to determine the motor position.

A classical approach to gaze controllers is to model several different controllers, each capable of a different function: for example, a specific controller for vergence, one for version, another for saccades, and so on. These controllers are referred to as *layered controllers*. The Harvard Head is an example of layered control ([23]): each subcontroller is responsible for a different movement. Another example of a layered controller is developed at the University of Rochester, named the Rochester Robot. It is designed for binocular gaze holding and is a good example of layered oculomotor control ([19], [27]). The robot requires 5 separate, independent controllers: saccadic, smooth pursuit, vergence, VOR, and head controller.

A hierarchical control scheme jointly developed by the University of Auckland, the Technical University of Berlin, and University of Erlangen-Nuremberg distributes various control loops to control a robotic head with ten degrees of freedom [89]. The controller hierarchy is modular: it seeks to exploit intrinsic competencies of lower level modules by higher level control strategies.

Other controllers do not necessarily use two cameras for the vision system. There are also implementations ([4], [30]) that use monocular tracking. On the other hand, the KTH [5] head, which is mounted on a navigational robot, needs peripheral vision to avoid collisions. Hence, it uses four cameras: two to model the eyes, and an extra two to implement peripheral vision.

3.1.2 Introduction to the Basic Ideas of the Controller

Twenty-five centuries ago, Aristotle, said that "the mathematical sciences particularly exhibit order, symmetry, and limitation; and these are the greatest forms of beauty." (Metaphysica, 3-1078b). Taking into account the way in which nature conflates beauty and functionality, Galiana and Outerbridge [35] presented an integrated model that implements the slow and fast phases of the vestibulo-ocular reflex (VOR) by means of a bilateral symmetrical controller.

Serving to stabilize the visual world during head movements [37], the VOR estimates the head velocity and converts it into eye position commands. The bilateral model of the VOR can reproduce biologically observed *ocular nystagmus*, alternating slow- and fast-phase eye movements in the same plane as the head movement, by changing the effective connections in the motor circuits shared by the fast and slow phase systems. The change is implemented via internal switching and hence the problem of instability due to state discontinuities is avoided. Instead, there is a seamless transition between slow phase and fast phase modes. Furthermore, no synchronization is required since all the controller modes share the same states.

An important characteristic of the VOR bilateral controller concerns the modeling of the coupling between corresponding networks on the two sides of the brain. Movements of the two eyes need to be coupled to insure fused binocular images. Coupling between the eye networks, as demonstrated by the brainstem-cerebellum



Figure 3-1: Bilateral controller conceptual diagram. The error signals are fused together at the input together with the Attention Cues. The motor command signals are received from the same controller.

symmetry, involves reciprocal inhibitory pathways that function as additional positive feedback loops. Galiana [37] describes that this coupling can perform important operational and adaptive functions that will be described in the following pages.

Part of the beauty of the bilateral controller is the superposition of sensory signals from the input, and motor drives from the output of the controller, as seen in Figure 3-1. This sensory-motor fusion of the signals resembles the physiological function of the vestibular nuclei.

By inference, the bilateral controller of the VOR can provide conjugate, vergence, pursuit, saccades and VOR by using one simple controller that imitates biology. This controller will serve as the basis for the binocular eye-head controller that drives the robot's gaze and thus, a more detailed analysis is appropriate.

3.1.3 Coordinate System

The coordinate system for horizontal eye movements is shown in Figure 3-2. As already stated, the bilateral controller geometry exploits symmetry. Hence, each eye's coordinate system is a mirror image of the other's in the horizontal plane. Throughout this thesis, eye movements towards the lateral side are defined as positive, and eye movements towards the center are defined as negative. Conjugate eye movements, and head movements towards the right are also defined as positive.

It can be geometrically proven that, using the given coordinate system, vergence (E_V) angles and conjugate (E_C) angles in the horizontal plane can be expressed with the help of the left and right eye positions according to (3.1) and (3.2) respectively.

$$E_{v} = E_{R} + E_{L}$$
(3.1)

$$E_{\rm C} = \frac{E_{\rm R} - E_{\rm L}}{2} \tag{3.2}$$



Figure 3-2: Eye coordinate system. Lateral angles are defined as positive, nasal angles as negative. E_L is the left eye position and E_R is the right eye position. The conjugate eye position (E_C) is the average direction in which the eyes are pointing, while the vergence eye position (E_V) is the angle between the left and right lines of sight. H represents head position. (Adapted from [91])

The significance of these two equations will become evident in the analysis of the Bilateral Controller that follows.

3.1.4 Bilateral Controller

The analysis of the bilateral gaze controller starts with the presentation of the butterfly that was first introduced by Galiana and Outerbridge [35]. Exploiting its symmetrical structure, it can coordinate simultaneously the vergence and conjugate components of pursuit eye movements in one dimension.

The butterfly resembles the operation of a differential amplifier since each one of the outputs can be expressed with the help of a difference –which corresponds to conjugate- and a common –which corresponds to vergence- mode. By direct inspection in Figure 3-3, the motor drives can be expressed with the help of the current inputs and current and previous outputs.

$$D_L(z) = i_L(z) + dM(z)D_L(z) - gD_R(z)$$
 (3.3)

$$D_R(z) = i_R(z) + dM(z)D_R(z) - gD_L(z)$$
 (3.4)

Where D_L and D_R are the butterfly outputs, g and d are gains and i_L and i_R are the controller's inputs, z being the Z-transform operator.

By solving (3.3) and (3.4) with respect to the output drives D_L and D_R we find the butterfly's system transfer functions:

$$D_{L}(z) = \frac{1}{2} \frac{(i_{R}(z) + i_{L}(z))}{(1 + g - dM(z))} - \frac{1}{2} \frac{(i_{R}(z) - i_{L}(z))}{(1 - g - dM(z))}$$
(3.5)

$$D_{R}(z) = \frac{1}{2} \frac{(i_{R}(z) + i_{L}(z))}{(1 + g - dM(z))} + \frac{1}{2} \frac{(i_{R}(z) - i_{L}(z))}{(1 - g - dM(z))}$$
(3.6)

Defining E_L and E_R to be the left and right eye directions respectively, we can formulate the overall transfer function of the system as :

$$E_{L,R}(z) = P_{L,R}(z) \cdot D_{L,R}(z)$$
(3.7)

where $P_{L,R}(z)$ is a model of the left and right eye plants. Since the controller is considered to be symmetric, from now on we consider that $P_L(z) = P_R(z)$.

The conjugate and the vergence modes of the motor drives are:

$$D_{conj}(z) = \frac{1}{2} \cdot (D_{R}(z) - D_{L}(z)) = \frac{1}{2} \cdot \frac{i_{R}(z) - i_{L}(z)}{1 - g - dM(z)}$$
(3.8)

$$D_{verg}(z) = D_R(z) + D_L(z) = \frac{i_R(z) + i_L(z)}{1 + g - dM(z)}$$
 (3.9)

Using the above definitions, equations (3.5) and (3.6) can be rewritten with respect to the vergence and conjugate mode of the motor drives as:



Figure 3-3: The symmetric controller is the central control unit of the bilateral controller that drives the eye-head system. i_R and i_L are the generalized bilateral inputs, D_L and D_R are the motor drives sent to the eye plants, g is the cross-commissural gain, M is a model of the eye plant dynamics and d is the gain of the unilateral component of the butterfly.

$$D_{L}(z) = \frac{1}{2} \cdot D_{verg}(z) - D_{conj}(z)$$
 (3.10)

$$D_{R}(z) = \frac{1}{2} \cdot D_{verg}(z) + D_{conj}(z)$$
(3.11)

Accordingly, and since we have considered that $P_L(z) = P_R(z)$ the eye positions can be separated geometrically as:

$$E_{conj}(z) = \frac{1}{2} \cdot (E_R(z) - E_L(z))$$
 (3.12)

$$E_{verg}(z) = E_R(z) + E_L(z)$$
 (3.13)

In Figure 3-3, M(z) is assumed to be simplified but accurate model in the operating frequency range of the eye plant P(z). In the literature [37], it is considered to be a 2nd order low-pass filter with a transfer function:

$$M(z) = \frac{(b_0 + b_1 z^{-1} + b_2 z^{-2})}{(1 + a_1 z^{-1} + a_2 z^{-2})}$$
(3.14)

3.1.5 Smooth Pursuit Mode

The butterfly operates in slow phase mode within a visual feedback loop as seen in Figure 3-4. In the presence of a selected target, the ideal command signals (T_L, T_R) are generated. The actual eye positions E_L and E_R are compared against the command signals in order to generate the left and the right retinal position errors, namely e_L and e_R . Following the error signal generation, a linear mapping process takes place which transforms the sensory signals to motor error signals. These motor signals constitute the inputs to the butterfly that has already been described.

Physiologically, there is evidence of a latency that exists between the appearance of a target and the activation of the ocular muscles due to visual perception that lasts approximately 150 ms. In the context of the implementation on the robot: A delay of approximately 66 ms is introduced into the loop: the frame grabbing and the blob^{*} detection processes require a not negligible time of 2 samples (frames) to be concluded.

As regards the mapping of the retinal errors to motor error signals, the conjugate and the vergence mode of the input signals $i_{R,L}$ is a linear combination of the conjugate and the vergence mode of the monocular error signals, respectively. Hence, the mapping functions K_{verg} and K_{conj} are dynamic functions which satisfy equations (3.15) and (3.16).

$$K_{verg}(z) = \frac{i_{verg}(z)}{e_{verg}(z)}$$
(3.15)

$$K_{conj}(z) = \frac{i_{conj}(z)}{e_{conj}(z)}$$
(3.16)

^{*} a group of adjacent pixels with almost equal intensity values that create an homogenous region in terms of brightness.



Figure 3-4: Smooth Pursuit Bilateral Controller. Vision provides a closed loop to the controller. When a target is identified, it is compared to the current position of the eyes and the error produced is mapped into motor signals i_L and i_R . A two sample delay is required to model the processing delay.

Equations (3.5) - (3.7) and (3.15) - (3.16) illustrate the attribute of the butterfly that its output is a combination of two linearly independent modes, namely the common mode and the difference mode. Explicitly, and assuming structural symmetry, the vergence mode of the output is a linear combination of the vergence mode of the input while the conjugate mode of the output is also a linear combination of the conjugate

input. That being said, a breakdown of the controller to two sub-controllers follows. The system transfer functions for these two independent controllers, with visual feedback from each eye, are:

$$\frac{E_{verg}(z)}{T_{verg}(z)} = \frac{K_{verg}(z) \cdot z^{-2} M(z)}{1 + g + (K_{verg}(z) \cdot z^{-2} - d) M(z)}$$
(3.17)

$$\frac{E_{conj}(z)}{T_{conj}(z)} = \frac{K_{conj}(z) \cdot z^{-2} M(z)}{1 - g + (K_{conj}(z) \cdot z^{-2} - d) M(z)}$$
(3.18)

where the parameter 'g' represents the gain due to the flow of the information across the midline from one side of the brainstem to the other, the parameter 'd' is the efference copy feedback gain and the parameters $K_{conj}(z)$ and $K_{verg}(z)$ control the controller's closed-loop tracking bandwidth.

Since the bandwidth of the vergence and the conjugate controller is limited by unavoidable delays, another process is required in order to cope with situations in which a fast change in gaze direction is needed. In the next section, the generation of the fast phase from the controller is discussed.

3.1.6 Saccade Mode

The requirement of orienting the foveae to a new focus of attention or following a target that moves faster than the maximum allowed pursuit movement as defined by the bandwidths $K_{conj}(z)$ and $K_{verg}(z)$ introduces the need for a brief ballistic jump of the



real world

Figure 3-5: Saccadic Controller. The cross-commissural links break apart and the visual feedback is sustained. The saccades are performed individual for each eye. The initial target T_s is stored at the beginning of the saccade, d_s is the saccade feedback inner loop gain.

eyes. In order to achieve higher bandwidths, the saccade circuit, shown in Figure 3-5, is blind to visual inputs outside the controller to avoid the limiting effects of visual delays.

Instead, an efference copy of the last estimate of the target location T_s , when the saccade mode takes over, is used as the location in space for the location goal. In other words, the saccade is a blind step response to an input that is equal to the last retinal error of the pursuit system added to the estimated eye position. In the previous robotic implementation, T_s was a snapshot at the time of the mode change. In the work to be described the saccade goal also makes use of prediction over the expected duration of a saccade as previously described in the theoretical work of Lee [60].

In addition, the projections across the midline are considered negligible meaning that each eye operates independently of the other in order to null its own gaze error. K_s controls the bandwidth of the saccadic controller. The transfer function of the saccadic system is given by equation (3.19), where d_s is the inner loop gain.

$$\frac{E_{L,R}(z)}{T_{L,R}(z)} = \frac{K_s(z) P(z)}{1 - d_s M(z) + K_s(z) P(z)}$$
(3.19)

When the motor error (e^{*}) becomes smaller than a certain threshold, the saccade mode terminates and the pursuit system takes charge again of the target tracking process.

3.1.7 Combining Fast and Slow Phases together

The two operating modes of the controller can be combined together in the same circuit as shown in Figure 3-6. The controller as shown is in pursuit mode. In order to operate in saccade mode, the gains "d" of the inner loops must switch to fast phase value (d_s) , the input to the sensory fusion point must disconnect from the visual signal and the cross-commissural gain (g) must be set to zero.

The decision mechanism that defines the switching strategy assumes that the initialization of the controller is always done in the pursuit mode. Hence, an error signal,



Figure 3-6: The 1-D controller that combines fast and slow phases. The controller as shown runs on pursuit mode. Running in the saccadic mode, the switches on the model are turned to "fast" position.

is computed during the first iteration of the controller and it is compared with a threshold θ . When the size of the error is larger than that of the threshold, the system switches to saccade mode. A return to the pursuit mode is decided only when the amplitude of the internal error signal ($e_{R,L}^*$) is less than a certain percentage (ρ) of the saccade amplitude Δ .

The sample and hold mechanism is used to hold the target position estimate of the last iteration of the slow phase mode throughout the fast phase operation. In other words,

the Sample and Hold box is in sample mode during pursuit and in hold mode during saccades.

3.1.8 Combining Eye and Head Positions: Gaze

In Figure 3-7 an upgrade of the bilateral model is shown in order to control the movement of a free head and coordinate this with binocular movements to control gaze. The gaze direction of each eye, G is defined as

$$G_{L,R}(z) = E_{L,R}(z) + H(z)$$
 (3.20)

where $E_{L,R}$ is the position of the eyes, H is the angle of the rotated head as described in Figure 3-2 and $z = e^{-j\omega}$. The rotation of the head does not affect the relative position of the two eyes. Therefore, only the conjugate movement of the gaze depends on the head rotation.

The system transfer functions (3.17) and (3.18) have to be modified to take into account the addition of the head. Since the vergence component of the movement is not affected by the head addition, it is expected that the vergence transfer function stays the same. Indeed,

$$\frac{E_{conj}}{T_{conj}} = \frac{K_{conj}(z) \cdot z^{-2} (P(z) + K_{Head} P_{H}(z))}{1 - g - dM(z) + K_{conj}(z) \cdot z^{-2} (P(z) + K_{Head} P_{H}(z))}$$
(3.21)

$$\frac{E_{verg}}{T_{verg}} = \frac{K_{verg}(z) \cdot z^{-2} P(z)}{1 + g - dM(z) + K_{verg}(z) \cdot z^{-2} P(z)}$$
(3.22)



Figure 3-7: A 1-D gaze model. A model of the head is integrated inside the current controller and an efference copy of the head position H^{*} is obtained. Sensory fusion is also introduced: The inertial input is superimposed to the motion commands. The VOR signal is added to the retinal error inputs of the butterfly.

3.1.9 Vertical Controller

The vertical controller uses the same notions as the horizontal controller that has already been presented. However, as defined in Figure 3-8 both eyes now have the same coordinate system that is identical to the one of the head. Due to this discrepancy, the horizontal common mode has become the difference mode in the vertical and the hori-



Figure 3-8: Ocular and Gaze Vertical Coordinate System. (modified from [91])

zontal difference mode has become the vertical common mode. Hence, the vertical gaze is calculated differently than the horizontal one. Also, it is important to note that the mapping of the retinal error to the sensory inputs has changed accordingly, as it is illustrated in equations (3.23) and (3.24).

$$i_{L} = K_{conj}(z)(e_{R} + e_{L}) - K_{verg}(z)(e_{R} - e_{L})$$
 (3.23)

$$i_{R} = K_{conj}(z)(e_{R} + e_{L}) + K_{verg}(z)(e_{R} - e_{L})$$
 (3.24)

3.1.10 Combination of Horizontal and Vertical Controllers

In Figure 3-9, a flow graph of a horizontal controller running in parallel with a vertical controller is presented. Note that the two controllers are not really independent since they are interconnected by a shared visual loop. Physiological findings dictate that both the horizontal and the vertical system must be in the same operating mode in any instant. Hence, a switching box is responsible to link the two controllers, thus forcing them to be in the same phase of movement. The decision mechanism is exactly the same that has already been described in page 34, but now the input data is kept in a 2-D variable that is composed from the inputs taken from both the controllers.



Figure 3-9: 2-D controller. The x-indexed variables refer to the horizontal dimensions while the yindexed variables to the vertical. The two bilateral controllers are linked together by a switch box that determines the mode in which both they will operate at the same time.

At this point, the literature review and the analysis of the theory of the bilateral controller is concluded. In the following pages of the chapter, a brief analysis is presented of the most important components used for the experimental implementation of the controller.

3.2 Description of the Experimental Implementation

In the remaining pages of this chapter, a brief presentation of the physical implementation of the robotic head is provided as the test-bed for the bilateral gaze controller.

- The *host computer* is an INTEL P4 1.6 GHz operating on a Windows 2000 platform and serves as the Central Processing Unit of the robotic system. The bilateral gaze controller, which is implemented in C under the Visual Studio C++ 6.0 SP5 environment, communicates with the robot using the motor drive commands and the functions on the frame grabber. Matrox Imaging Library (MIL) 7.5 is also used to provide the Visual Studio Environment with useful and fast imaging processing functions. National Instruments NI-DAQ 6.9.2 provides the necessary functions, libraries and drivers to use the A/D and D/A cards with C inside the windows environment. MATLAB 7.5 is also used to further analyze the acquired experimental data.
- Two ELMO CCD MN401E *cameras*, which provide an NTSC signal to the frame grabber at 30 frames per second, model the system's stereo vision. Each one of the two cameras has two degrees of freedom (vertical and horizontal movement). Since the cameras must be synchronized, the NTSC signal from the right camera, which is the master, is connected to the external sync of the left camera the slave. The output analog video signal conforms to the NTSC color standard which is a 525 lines, 60 fields per second, 2:1 interlaced system that uses YIQ color space -Y is the black and white portion of the image, with the I and Q parts forming the color components.
- A MATROX METEOR II / Multichannel *frame grabber*, which digitizes the NTSC signal into a 2D buffer for further processing and analysis. The digitized image has a resolution of 640 pixels by 480 pixels. The frame grabber performs on board computations using the MIL software package and provides target blob localization. Concerning the sampling rate, it synchronizes the system by providing a clock signal

based on a frame grabbing rate of 30 fps. In order to obtain both left and right images at the same time, both cameras must be sampled on the same channel: to accomplish this, the frame grabber is tricked into believing that there is only one camera connected by assigning a different color band to each camera, namely the left camera is connected to the blue (B) band, and the right camera is connected to the red (R) band. Hence, visual feedback from each camera will be processed based on intensity signals.

- The *neck* provides yaw and pitch motion for the robotic head and is controlled by velocity servo motors with two degrees of freedom (vertical and horizontal movement with motor velocity proportional to the input voltage). A combination of the two motors, an inner and outer motor, provides the required movements. Activating only the outer motor produces pure pitch while activating both the inner and the outer motor with the same drive produces a pure yaw movement from the 'neck'.
- A Watson ARS-C141 angular velocity sensor is mounted parallel to the head's yaw axis of rotation and measures angular head velocity around the horizontal axis. It is used to sense head perturbations and the head's position in 1-D space. The sensor consists of piezoelectric bending elements that are mounted parallel to each other to a rigid base in a tuning fork configuration. The out-of-plane Corriolis force caused by rotation can then be demodulated to provide the measured angular velocity over a range of ±600 deg/s with linearity error lower than 0.1%. A 6th order low-pass Bessel filter ensures that no aliasing occurs on the sensor signal when sampled at the frame rate (30Hz). The filter cutoff is adjusted by an appropriate choice of resistors so that the visual and inertial sensors can be sampled at the same rate. The chosen cutoff frequency is ~10 Hz, which is less than the Nyquist frequency (15 Hz.)
- The National Instruments PCI 6062E *Analog-to-Digital Converter* performs the ADC of the analog velocity signals and provides a clock signal for analog acquisition.



Figure 3-10: Block diagram of the robotic implementation. The system has a visual input (cameras), an inertial input (angular velocity sensor), four outputs that drive the PWM for the camera servo motors and two outputs that drive the velocity controlled neck. (Adapted from [91])

- The National Instruments PCI 6703 *Digital-to-Analog Converter* translates digital voltage to analog DC values used to drive the motors.
- The four *Digital Servo motors* (two of the type JRDS3421 and two of the type JRDS8411SA) provide each camera with yaw and pitch movement respectively and constitute the physiological counterpart of the muscles that move the eyes. The position-controlled servo motors for the cameras require three signals to operate: i) Vcc, the 5 Volt input, ii) Gnd, the 0 Volt ground, and iii) the pulse width modulation (PWM) encoded signal providing the desired position. The servo that permits vertical motion was chosen to provide more torque (155 oz/in) than the one providing

horizontal motion (65 oz/in) since the second motor needs only support the weight of the camera, whereas the first must support both the camera and the horizontal servo.

• The Pulse Width Modulation Module TL494 converts DC voltage into PWM signals in order to drive the four Digital Servo motors discussed before. The camera servo motors used in the robotic head can operate with a PWM frequency between ~50 Hz and ~300 Hz. The faster the frequency, the more often the error correction is performed: this improves holding power and disturbance correction. Hence, the designed PWM frequency was chosen to be around 300Hz limiting the pulse period to approximately 3.3ms. By definition, a servo is at its neutral point, or center, for a pulse width of 1.5ms. A 1 ms pulse forces the servo to its leftmost position while a 2 ms pulse forces it to its rightmost position.

The next chapter will illustrate the performance of this robotic vision system with alternate control strategies.

Chapter 4 Adding Prediction and Memory in the Biomimetic Controller

4.1. Starting Point: Pursuit Controller with Positional Information

As described in Chapter 2, the photoreceptor mosaic of the vertebrate retina transduces light energy in the form of photons into neural activity, ultimately in the form of action potentials. The spatial resolution of this transduction system is limited by the resolution of the photoreceptor mosaic, but only if the eye can be kept stationary with regard to the objects in the external world that are subjects of visual analysis. (for the moment we ignore micro-saccades that are believed to be necessary to preserve functional vision in a system with high-pass properties). Thus, stabilizing the retina on moving or stationary targets is a critical challenge to effective vision. In this chapter the neural and behavioral systems vertebrates use to achieve effective retinal stabilization is examined and their implementation in the robotic head is described and further discussed.

In Figure 3-9, the input to the controller is retinal error which depends solely on the fovea's positional deviation from the visual goal. The sensory information is mapped into the motor drive using the functions already described in section 3.1.4 and reproduced in equations (4.1) and (4.2) for convenience.

$$K_{verg}(z) = \frac{i_{verg}(z)}{e_{verg}(z)}$$
(4.1)

$$K_{conj}(z) = \frac{i_{conj}(z)}{e_{conj}(z)}$$
(4.2)

where $e_{verg}(e_{conj})$ is the visually derived vergence (conjugate) error and $i_{verg}(i_{conj})$ is the vergence (conjugate) motor error signal.

The starting point of this chapter is these mapping functions first described in [91]. Being scalars, they deprive the slow phase controller of memory and prediction and only allow the comparison of the absolute value of the positional error with a predefined threshold. That threshold is used to decide upon initiating a fast phase movement that stops when the retinal error decreases to a certain percentage of the saccadic goal, derived from the positional error right before the saccade commences.

With only the retinal error as guidance, the controller never stops correcting itself with saccades: In order to follow a target trajectory, it will need to alternate between pursuits and saccades comparing continuously the error with a lower and an upper bandwidth respectively. Namely, the gaze trajectory depends on switching between the fast and the slow phase to compensate for the lack of knowledge of the target speed.

The nystagmus in the controller's output is independent of the speed of the target since it is the non-ideal gain of the pursuit mode that makes the catch-up saccades necessary. The gaze is never stabilized on the target and the main objective of the slow pursuit system, which is the centralization of the target on the fovea, is never achieved. The system can never maintain perfect tracking, since that would reduce the retinal error driving the system to zero, and therefore transiently reduce eye velocity to zero. Consequently, this control approach is clearly inadequate and improvements towards the direction of matching the target and the gaze velocities should be made.

4.2. Adding Prediction and History Information on the Pursuit Controller

Going back to the heart of the problem, the incessant need for switching and its timing, two directions should be followed: minimize the velocity disharmony between the target and the gaze trajectory and maximize the gain of the pursuit controller output so that the trajectory curve lies within the foveal range from the target for as much time as possible. By providing the first improvement, the controller postpones the need for saccades allowing the system to work in pursuit mode for longer periods. The second proposed amelioration ideally obliterates the need for saccades since the aim of the controller, positioning the target on the fovea, can be fulfilled without them, at least over a desired bandwidth.

4.2.1 Slip Signals in Physiology

From a neurophysiological and anatomical point of view, the middle temporal (MT) area and medial superior temporal (MST) area seem to be intimately involved in smooth pursuit. Moreover, neurophysiological results ([32]) support the idea that the major computation for the smooth pursuit maintenance phase is performed in the MST area.

In humans and most other vertebrates, the retina projects directly to a midbrain area just rostral to the superior colliculus called the pretectum. Many neurons in this area become active when the visual world slips in a particular direction, and their firing rate increases (up to a point). Thus, pretectal neurons encode the velocity and direction of retinal slip [42]. They project directly to the vestibular nuclei via pontine and medullary relays. In fact, many vestibular neurons can be activated by both vestibular and visual stimuli [41], suggesting that the two types of stimuli access a common circuitry for eye velocity control.

Two hypotheses have been proposed to explain the pursuit drive. One hypothesis states that the speed and direction of slip are the only visually derived inputs used by the smooth pursuit system [84]. The other hypothesis states that information about the position, speed, and acceleration of the target is combined to compute a dynamic force structure optimized for the properties of a small visual target that moves smoothly across its background [62]. The first hypothesis uses a minimum of visual information and structures the pursuit drive based entirely on the properties of the target's velocity. Characteristically, it states that the response to a step is not due to a position error but to a misinterpretation of the step as a velocity impulse. The second hypothesis structures the pursuit drive by analyzing the visual target's movement and the different kind of errors (position, speed, acceleration) are added linearly. These two hypotheses place the

computational burden on different brain structures: the first hypothesis places the burden in the oculomotor brainstem whereas the latter hypothesis places it in the largely cortical areas that compute information related to target motion. For the second hypothesis, important areas include as well the striate cortex and the middle temporal visual area, and probably the medial superior temporal visual area and the posterior parietal cortex.

Roughly in the area that is called MST, a group of neurons was found [49] that responds to gaze velocity, i.e. the velocity of the gaze in the spatial coordinates. This result was supported by experiments in which the activities of the same neurons were very similar in either pure eye tracking of a sinusoid or during vestibulo-ocular reflex (VOR) cancellation, where the target was a sinusoid and the head of the monkey was also externally moved sinusoidally, requiring the monkey to keep its eyes still and cancel the VOR in order to keep its gaze on the target. On the other hand, the same neurons had no discharge properties when the monkey was oscillated sinusoidally in the dark (a pure VOR task). Therefore, activity in these visual tracking neurons seems to originate from highly preprocessed visual inputs and not from efferent or afferent signals that are related to eye movement.

The dorsolateral pontine nucleus (DLPN) is a critical link in the smooth pursuit system, bridging between the cortical motion-processing systems and the oculomotor portions of the cerebellum and brainstem. The DLPN contains neurons that encode the direction and speed of pursuit, the direction and speed of target motion, or both[95]. Thus, the DLPN could, in principle, process information about both the sensory representations of targets and the motor error signals required to drive the eyes. The output of the DLPN passes to the cerebellum (the flocculus, paraflocculus, and vermis), where there are neurons whose firing rate is tightly coupled to the velocity of eye rotation specifically during smooth pursuit. These neurons in turn project to the vestibular nuclei, where their error signals are thought to be integrated into the vestibulopontine oculomotor systems [59].

The identified brainstem pursuit pathways form a portion of the complete pursuit system. Signals that indicate target motion are critical to their function. This information appears to be provided to the pursuit system by the cortical motion system including areas MT and MST. These areas compute the direction and speed of moving stimuli throughout the visual field and pass this information to the DLPN, directly and via the posterior parietal cortex and the frontal eye fields.

Slow eye movements can be induced by either vestibular or visual stimuli. Stimulus velocity has been shown to be the primary signal driving the oculomotor system ([25], [26], [65], [78]). Slow eye movements generated by the vestibular system are essentially a response to head velocity ([65], [93]). Thus the signal they generate is proportional to head velocity rather than to head acceleration [65]. Rashbass [77] was the first to show that the control system mediating visually guided smooth pursuit is primarily responsive to target velocity.

It was shown that eye velocity can change during pursuit with latencies as short as 75 msec [78]. This indicates that target motion is probably monitored continuously rather than intermittently ([8], [77]). In addition, there is a high forward gain in the input-output relationship during pursuit. This is shown by the exact matching of eye velocity to target velocity at low speeds [111].

Using Helmholtz's concept of efference copy [39], Young, Forster & Van Houtte [110] assumed that a signal representing eye velocity was fed back from the oculomotor to the visual system. When combined with the retinal error signal, it gave a "perceived" target velocity, which then drove the oculomotor system. This supposition avoided the instability of the original model and accounted for its high open-loop gain. Yasui et al. [109] provided evidence for the "perceptual feedback" hypothesis. They recorded nystagmus induced by angular rotation in darkness with a foveal after-image. Although there was no retinal error in either case, compensatory eye movements were improved by the after-image. They suggested that the difference was due to a perceptual feedback loop that had been activated by the after-image.

In summary, target velocity is postulated as the major driving signal for smooth pursuit. It is presumably formed by a combination of retinal error velocity (slip) and an efferent signal that is fed back to the visual system from the oculomotor system. The slow eye movement control system is capable of predictive and selective following of targets.

4.2.2 Integration in Physiology

In general, neurobiologists have tended to view the pursuit system as follows: retinal slip, restricted to a selected portion of the visual world, is minimized by adjusting eye velocity until slip is reduced or eliminated. Knowledge about the speed and direction of the target's motion across the retina and the eye's current movement is used to compute a desired speed and direction of eye movement. The eye velocity signal is then passed via the brainstem oculomotor nuclei to the extraocular muscles to control the dynamic movement of the eye. This dynamic signal, like the dynamic signals for other eye movements, is presumed to be integrated to compute the static signal necessary to maintain the eye in a fixed position should the eye stop. This integrator is thought to reside in the prepositus nucleus of the hypoglossal nucleus.

The presence of an integrator in oculomotor pathways has been inferred from stimulation studies ([24], [50]) and from studies on the vestibulo-ocular reflex arc ([79], [80], [93], [95]). This integrator has a time constant longer than 20 sec in human [12]. However, after cerebellar lesions its time constant can fall to 1 sec [20]. The motoneuron by itself is probably not capable of such integration, since it needs continuous excitation to fire at steady rates [88]. Nor is this integration likely to take place on the eye muscles, because activity related to position of the eyes is already present in motoneurons. Therefore it must occur centrally [93].

Krauzlis et al [53] found that the SC neurons encode retinotopically mapped motor errors, which are shared by both fast and slow eye movements. As described in Chapter 2, when a visual scene contains multiple potential targets, the vision system chooses a target that is the vector average of different attributes. The various attributes are superimposed as activity levels located on the collicular motor map [38]. Also important is the way errors are coded by SC neurons. A typical neural pattern on the collicular map is not a single point impulse function, but rather a hill which involves many neurons. The motor error is encoded by taking the population average of the hills. Motor errors are coded in retinal coordinates [51] relative to the foveal area. During any slow pursuit gaze segment, the activity hill moves towards the origin[†], and any point *P* in the hill is represented by different SC neurons at different times. Hence, the location of *P* is a function of time with respect to the SC coordinates (x, y), and can be written as (X(t), Y(t)). Since smooth pursuit is relatively sluggish (hence earning the name "slow phase"), each *P* will have some displacement during the smooth pursuit preparation, and the neuronal outputs have to be integrated over the interval during which this displacement is made ([13], [14], [15], [60]).

4.3. Proportional – Derivative – Integral (PDI) Analysis

Smooth pursuit has traditionally been described as negative feedback control system ([35], [79]). In Sejean's work [91], the feedback signal was a function of the position error. In the current thesis, the speed of that error as well as its past values will contribute to the error signal needed to be minimized.

The stimulus to the pursuit system is now a function of the velocity of motion, or slip, of the visual image of the target as it moves away from the fovea, across the retina. The retinal error velocity is defined as the difference between target velocity and gaze velocity and it reflects the physical fact that the retina is attached to the eye. This retinal error is the signal used by the system to generate an eye velocity command. It is amplified by the brain to generate an eye movement that will catch up with the target.

In the discrete domain, the error velocity is approximated by the first derivative of the error signal, as described in equation

$$RS = \dot{T} - \dot{G} = \dot{e} \cong \frac{e[n] - e[n-1]}{T_s}$$
 (4.3)

[†] Movement of an activity hill in the SC is still a controversial issue in literature. It is better accepted and adequately verified in cat and not in monkey map. For the scope of this thesis, the assumption that each point P will have some displacement during slow phase mode is followed.



Figure 4-1: 1-D PDI Conjugate Pursuit Controller. The delayed retinal error is multiplied by a constant K_{conj} , differentiated and then multiplied by a different constant K_{conj}^{slip} , integrated and then multiplied by a third constant K_{conj}^{int} . The three signals are superimposed in the summing junction and provide the motor drive i_c for the controller.



Figure 4-2: 1-D PDI Vergence Pursuit Controller. The delayed retinal error is multiplied by a constant K_{verg} , differentiated and then multiplied by a different constant K_{verg}^{slip} , integrated and then multiplied by a third constant K_{verg}^{int} . The three signals are superimposed in the summing junction and provide the motor drive i_V for the controller.

The stimulus to pursuit movement is also a function of the past values of the position error. Given a function f, the function g is named the *integral* of f if and only if

 $g(t) = \int_{t} f(t)dt$. The Z- transform of an integrator is approximated by the equation

$$G(z) = F(z)\frac{T_s}{z-1}$$
(4.4)

where G(z) and F(z) are the Z-transforms of g(t) and f(t) respectively and T_s is the sampling time.

The addition of the slip and the integration signals on the controller should only affect the mapping function, meaning any contribution that they have is added to the sensory fusion node. Following the conjugate and vergence eye movement convention introduced in equations (3.17) and (3.18), the PDI mapping functions described in general in equations (3.15) and (3.16) can be expressed analytically in equations (4.5) and (4.6) that follow.

$$K_{conj}(z) = K_{conj} + (1 - z^{-1}) \cdot K_{conj}^{slip} \cdot F + K_{conj}^{int} \cdot T_S / (z - 1)$$
(4.5)

$$K_{verg}(z) = K_{verg} + (1 - z^{-1}) \cdot K_{verg}^{slip} \cdot F + K_{verg}^{int} \cdot T_S / (z - 1)$$
(4.6)

The conjugate and vergence PDI pursuit controllers with the enhanced mapping functions are shown in Figure 4-1 and Figure 4-2 respectively. The transfer functions can now be written analytically as follows:

$$\frac{G_{conj}(z)}{T_{conj}(z)} = \frac{2z^{-2}(K_{conj} + (1 - z^{-1})K_{conj}^{slip}F + K_{conj}^{int}T_S/(z - 1))[K_{head}P_H(z) + P(z)]}{1 - g - dM(z) + \frac{2}{z^2}(K_{conj} + (1 - z^{-1})K_{conj}^{slip}F + \frac{K_{conj}^{int}T_S}{z - 1})[K_{head}P_H(z) + P(z)]}$$
(4.7)

$$\frac{G_{verg}(z)}{T_{verg}(z)} = \frac{2z^{-2}[K_{verg} + (1 - z^{-1})K_{verg}^{slip}F + K_{verg}^{int}T_S/(z - 1)]P(z)}{1 + g - dM(z) + \frac{2}{z^2}[K_{verg} + (1 - z^{-1})K_{verg}^{slip}F + \frac{K_{verg}^{int}T_S}{z - 1}]P(z)}$$
(4.8)

4.3 Effects of the PDI signal on the saccade controller and improvements on the switching strategy

The saccadic system brings in higher brainstem and cerebral mechanisms to identify a target and move gaze to it. The eye movements themselves are very fast, and they minimize the time lost in visual contact while gaze is shifting to a new target. However, the *saccadic latency* which is the time from the presentation of the target to the commencement of the saccade has been measured to be around 180 - 250 ms. These transmission times in the afferent and efferent pathways and the central processing time are likely to be appreciable due to the complexity of the pathway and contribute to the

retinal error during nystagmus. In the following, retinal slip is also considered to play an important role in the programming of the saccade *amplitude*.

4.4.1 Physiological Basis

Saccades are particularly frequent during pursuit with a low gain or following unexpected changes in velocity or direction of the target. The goal of this paragraph is to present the sensory conditions leading to the occurrence of catch-up saccades during smooth pursuit. Sensory cues available are position error and retinal slip, which are known to play different roles in saccades and smooth pursuit. An important property of saccadic generation is prediction ([29], [105]). When responding to a periodic target, the eyes can follow with zero phase (no delay) or can even lead the target.

The complex task of decision between two different control strategies, i.e., saccadic and smooth eye movements, gives insight concerning the way the CNS combines different control strategies to achieve a common goal. Indeed, understanding what are the sensory parameters that determine the triggering of catch-up saccades is fundamental when studying visual tracking of moving targets. Brouwer et al. [18] showed that it was possible to control the probability of occurrence of catch-up saccades by combining appropriate values of position and velocity steps in the target trajectory.

There are other examples of close interactions between the saccadic and smooth pursuit systems that have been recently described in the literature. It has been shown that there is a common motor or position error signal in the superior colliculus that could be shared by the saccadic and smooth pursuit systems ([10], [53]). At the premotor level, Missal et al. [68] have found neurons in the mesencephalon that are active during both kinds of eye movements, raising the possibility that they play a role in the synergy between saccadic and smooth pursuit systems. Another site of interaction between both systems is the oculomotor cerebellar vermis where lesions affect both saccades and smooth pursuit [96]. Krauzlis and Miles [54] reported that the vermis might contribute to the elaboration of an error signal that is common to both types of eye movements. Electrical stimulation in that structure evokes saccades or smooth eye movements, depending on the sensory context at the time of stimulation. Brower et al. [18] found that the sensory context could be described by the combination of Position Error (PE) and Retinal Slip (RS). Finally, at the cortical level, there is anatomical evidence for connections between structures containing subregions for saccades and pursuit [98].

Brouwer et al. [18] also discovered that the main parameter controlling the decision mechanism is the eye crossing time (the expected time in which the eye trajectory intercepts the target trajectory), which depends on two sensory signals: position error and retinal slip. In fact, catch-up saccades are triggered with a latency of nearly 100 ms when the value of the eye crossing time enters the saccade zone, which means that it is likely not possible to catch the target solely by means of a purely smooth acceleration.

4.4.2 Saccadic Prediction

In Sejean's implementation of the model [91], the saccadic controller does not rely on the effects of the derivative or the integral of the retinal error may have on gaze. Hence, the saccade goal is computed from a position error signal, relying on the internal efference copy eye position and the delayed retinal error.

In order to exploit the information in target velocity for the programming of saccade amplitudes, the retinal slip is calculated from an internal memory of retinal error. In biology, this information may be tapped directly from cells in the cerebellum or from the mossy fibers of the flocculus. On the PDI controller, the saccade amplitude (A_{SAC}) is no longer equal to the zero-order-hold sample of the retinal error. Instead, it is calculated using the mathematical expression:

$$A_{SAC} = e_{S_H} + R_S^* \cdot \tau_{SAC}$$
(4.9)

where $R_{S}^{*} = \frac{de_{S_{H}}}{dt}$ is the slip and τ_{SAC} is the weight on R_{s}^{*} in the saccade amplitude drive and designed as a time period, to conform to the units in equation (4.9). It is also related to the duration over which error velocity is integrated to provide the corrective value. Experimental simulations [60], defined an appropriate choice $\tau_{sac} = 3 \cdot T_s$ which agrees with the physiological approximations of nearly 100 ms for the saccade latency reported in the literature.

4.4.3 Revised Switching Strategy

Since slow phase behavior now depends on both retinal slip and retinal error, the decision for whether a transition between slow and fast phase should be made, should exploit both of these signals. From an engineering point of view, the decision threshold should carry an information for the failure (or success) of the slow pursuit gaze to predict and follow the velocity of the moving target.

That decision was made in [91] using only the current retinal error in a way that a saccade was triggered only if the retinal error was larger than a predefined threshold. Now, the threshold is dynamic by superimposing the weighted speed of change of the error. This composite quantity, expressed in equation 4.9, calculates for every instance the threshold ϵ in the strategy described in section 3.1.6.

$$\varepsilon = e_{COMP} = e + \frac{de}{dt} \cdot \tau_{COMP}$$
(4.10)

where e_{COMP} is the composite control signal, *e* is the visual (retinal) error, and τ_{COMP} is the constant of proportionality for the slip component. The threshold ' θ ' to enable saccades needs to be augmented to compensate for the use of a composite error signal.

4.4.4 Reducing Switching Noise

The proposed changes so far, for both fast and slow phase controls, lie well within what is supported by physiological research findings. In order to avoid any secondary artifacts created by their introduction in the controller, some non-linear actions need to be implemented. These actions are described in the remainder of this section.

4.4.4.1 Velocity Threshold to limit Saccade Duration

As described in page 33, on the initiation of a saccade, the last sample of the visual error is frozen and the desired amplitude Δ of the saccade is calculated using that value. Saccades, being essentially step responses, will have trajectories that reach a plateau in which the velocity of the gaze is near zero.

Referring to the switching scheme described in section 3.1.6, the strictness of the saccade termination decision is defined by the quotient ρ . Depending on how small a value of ρ has been predefined, the decision upon switching back to the pursuit controller, may have implications on the discrepancy between the target and the gaze trajectories. That is because the saccade amplitude may reach the switching threshold with a low speed, delaying the time the controller may directly view the target again, from the time the saccade has been initiated until the switching criterion has been satisfied. It is easy to understand that, if that is the case, then the whole purpose of having a saccade is defeated.

Hence, the saccadic controller now monitors its internal estimate of the gaze velocity (V_G) and when it becomes lower than a velocity threshold, namely θ_V , the saccade is turned off regardless of whether the switching criterion (involving the percentage of the frozen combined motor error) has been reached or not. The controller is then forced into slow phase so that the visual information can be updated.

4.4.4.2 Nullifying Retinal Slip (R_s) at the onset of the smooth pursuit segment

As it is described in Section 0, the motor drives to the controller depend on the speed of the retinal error. At the end of a saccade, the retinal slip is very high, leading to a temporarily large R_S value that can be considered as high energy noise when the positional error is small. Recalling that the visual input used is always delayed by two samples, one can understand that the position and velocity drives for the slow phase control go into conflict with each other. This effect is due to the fact that the slip component during mode transitions is high, while the gaze can still be near target.

For these reasons and in order to guard the stability of the system, it is beneficial to leave out sensory transients right after an internal mode switch. Hence, R_s is forced to become zero at the onset (first two samples) of a smooth pursuit segment, so that no transient perturbations, due to the large slip at the end of the saccade, will corrupt the gaze trajectory.

4.4.4.3 Saturation of the Retinal Slip during discontinuities

In cases where the target trajectory is not piece-wise continuous (target jumps), discontinuity spikes are introduced in the estimated target and retinal error accelerations. Due to the fact that in these situations the R_S value can be very high, a need for a nonlinear saturation in case of high slip values arises. It is worth mentioning that this property is also observed in biology [17]. Hence, the R_S^* value, that has been presented in section 4.4.2, is meaningless noise and its contribution to the programming of the saccade amplitude, as expressed mathematically in equation (4.9), is erroneous. R_S^* gives indeed valuable information about the current speed of the target, but only after the momentary instability has passed.

In order to identify the circumstances in which the saturation is needed, the acceleration of the retinal error is calculated and compared with a threshold ζ . A typical value of ζ is 5,000 pixels / s².

An alternative way of eliminating R_s^* spikes at target discontinuities is to apply a low-pass filter to the R_s^* signal, so that slow changes in R_s^* values will get through even at high levels, whereas high frequency spikes will be attenuated.

4.4 Final Controller Parameters and Setup Steps for Data Acquisition

The experiments that will be presented in the following pages of the chapter have been done on the system that has been described in Section 3.2. The running parameters for the P Controller results have been adapted from [91] and have been slightly modified. The running parameters for the PDI controller results were adapted from the simulation study results that have been presented in [60] in order to also compensate for the system dynamics that were not modelled. The scope of this chapter is to describe the implementation of the PDI controller on the robotic head and to show its significant contribution to the gaze. In Appendix 2 a method for observing the contribution of each parameter to the length and the flatness of the bandwidth as well as an optimal parameter set are presented. All the experiments presented in Section 4.6 have been conducted using the parameters presented in the following tables.

Parameters	Slow phase value	Fast phase value
g	0.001	-
d	(1-g)/1597.7	-
ds		1/1597.7
K _{conj}	2.65e-5	-
K _{verg}	2.65e-5	-
Ks		2.75e-4
K _{Head}	3	5
Tf_on (θ)/ Tf_off (ρ)	$\theta = 20$ (pixels)	ρ = 0.2
Refractory time	$T_{REFRACT} = 2.5 T_{S} (\sim 80 ms)$	$T_{SAC_REFRACT} = 3 T_{S} (\sim 100 ms)$

Table 4-1: Table of parameters used for the horizontal OCS controller in the P controller(see Figure 3-7 and Section 3.1.6).

Parameters	Slow phase value	Fast phase value
g	0.001	-
d	(1-g)/1584	-
d _s	-	1/1584
K _{conj}	2.2e-5	
K _{verg}	2.2e-5	-
Ks	-	2e-4
K _{Head}	3	5
Tf_on (θ)/ Tf_off (ρ)	$\theta = 20$ (pixels)	ρ = 0.2
Refractory time	$T_{REFRACT} = 2.5 T_{S} (\sim 80 ms)$	$T_{SAC_REFRACT} = 3 T_{S} (\sim 100 ms)$

 Table 4-2: Table of parameters used for the vertical OCS controller in the P controller (see Figure 3-7 and Section 3.1.6).

Parameters	Slow phase value	Fast phase value
g	0.001	-
d	(1-g)/1600.5	-
ds	-	1/1742.4
$\left(\mathrm{K_{conj}, K_{conj}^{slip}, K_{conj}^{int}} ight)$	$(4e^{-5}, -0.5e^{-6}, 3e^{-5})$	
$\left(K_{verg}, K_{verg}^{slip}, K_{verg}^{int}\right)$	(3.05e-5, -0.6 e-6, 1.15 e-5)	-
Ks	-	1.40e-4
K _{Head}	1.05	5
Tf_on (θ)/ Tf_off (ρ)	$\theta = 20$ (pixels)	ρ = 0.2
Refractory times	$T_{\text{REFRACT}} = 2.5 \text{ T}_{\text{S}} (\sim 80 \text{ ms})$	$T_{Sac_REFRACT} = 3 T_{S} (\sim 100 ms)$
θν	-	300 (pix/sec)
τ_{SAC}/τ_{COMP}	$\tau_{\rm COMP} = 0.5 {\rm Ts} (\sim 15 {\rm ms})$	$\tau_{SAC} = 3 \text{ Ts} (\sim 100 \text{ ms})$
clamp (ξ)	200 (pixels/s)	200 (pixels/s)
aclamp (ζ)	5000 (pixels/s ²)	

 Table 4-3: Table of parameters used for the horizontal OCS controller in the PDI controller (see Figure 4-1, Figure 4-2, Section 3.1.6 and Section 4.4.4).

Parameters	Slow phase value	Fast phase value
g	0.001	-
d	(1-g)/1690.1	-
ds	-	1/1690.1
$\left(\mathrm{K_{conj}, K_{conj}^{slip}, K_{conj}^{int}}\right)$	$(3e^{-5}, -0.78e^{-6}, 1.85e^{-5})$	-
$\left(\mathrm{K}_{\mathrm{verg}},\mathrm{K}_{\mathrm{verg}}^{\mathrm{slip}},\mathrm{K}_{\mathrm{verg}}^{\mathrm{int}} ight)$	(3 e-5, -0.7 e-6, 1.15 e-5)	-
Ks	-	1.5e-4
K _{Head}	2	3
Tf_on (θ)/ Tf_off (ρ)	$\theta = 20$ (pixels)	ρ = 0.2
Refractory time	$T_{REFRACT} = 2.5 T_{S} (\sim 80 ms)$	$T_{SAC_REFRACT} = 3 T_{S} (\sim 100 ms)$
θ_V	-	300 (pix/sec)
τ_{SAC}/τ_{COMP}	$\tau_{\rm COMP} = 0.5 {\rm Ts} (\sim 15 {\rm ms})$	Not used
Clamp (ξ)/ aclamp (ζ)	Not used	

 Table 4-4: Table of parameters used for the vertical OCS controller in the PDI controller (see Figure 4-1, Figure 4-2, Section 3.1.6 and Section 4.4.4).

 Concerning the procedural steps that were followed in order to acquire the data presented in Section 4.6, the protocol is described below.

Phase 1: Calibrate at the center.

- 1. The laser dot is positioned on the wall between the two eyes at the same height, to ensure that it can be easily localized.
- 2. The eyes are positioned looking straight up before the initiation of the experiment.
- 3. The controller is run once with pre-calculated centralized initial eye positions: the eyes make a small vergence movement in order to focus on the central laser dot and stay still.
- 4. The controller centers the eyes on the target driving the retinal error to zero.
- 5. The controller is turned off.
- 6. The controller's final states are saved and stored for both eyes.

Phase 2: Main Experiment.

- 1. The controller loads the initial conditions: As the origin of the experiment's space the initial position of the target is considered and not the position of the eyes.
- 2. The target moves: The laser dot executes a pre-programmed trajectory or is driven from the external function generators.
- 3. The head position, the left and right retinal errors, the left and right eye position estimates and other secondary signals (motor signals, saccade flags etc.) are recorded as well as the volts that drive the laser dot either from the external generators or from the pc analog output port (see Appendix 1).

The initialization was executed before acquiring data from the controller. The experimental data have been analyzed offline using MATLAB. In the following part of the thesis, the terms "eye" and "camera" will be used interchangeably.
4.5 Experimental Results[‡]

In this section, the experimental results are presented. The experiments are categorized into 3 categories, namely slow phase eye movements, eye movements with switching allowed and gaze response of the robotic head. In all categories, a 2D input is also used to verify that the controller works in both the horizontal and vertical planes.

4.6.1 Slow Phase

The analysis will begin by discussing the controller's properties in the frequency domain such as steady-state gain and bandwidth with the chosen parameters. An array of inputs illustrating the controller's response is then used to test the system: these consist of various frequency sinusoids and step inputs. In addition, the effect of the frequency of the input signal to the phase lag introduced to the controller's output is discussed.



Figure 4-3: 1st row: Horizontal Response of the P & PDI controllers to a chirp signal. Upper left is the conjugate eye movement (black thin line) of the P controller to a 0.1 Hz sinusoid target (thick gray line)[§]. Upper right is the conjugate eye movement (black thin line) of the P controller to a 0.1 Hz sinusoid target (thick gray line). 2nd row: Input laser signal, a sinusoid increasing in frequency from 0.1 Hz to 2 Hz within 90 sec. 3rd row: The conjugate Eye Response of the P controller. Black straight line: 3db point. 4th row: The conjugate Eye Response of the PDI controller.

[‡] The experimental results shown use the pixels-to-degrees and voltages-to-degrees mapping functions presented in Appendix A.

[§] The thickness of the cyan line (target trajectory) is equal to the typical size of the fovea.

4.6.1.1. Controller's Experimental Bandwidth - Head Fixed

In order to estimate the amelioration of the bandwidth of the robotic system that is available from the PDI controller, a chirp signal with increased amplitude linearly from 0.1 Hz to 2 Hz in 90 s was used as an input to the system and the conjugate eye position was computed from binocular retinal error and plotted. The experiments were done separately for the two dimensions and the results are presented in Figure 4-3 and Figure 4-4. From the observation of both figures, it is clear that both controllers work as low pass filters since they show the magnitude attenuation characteristic of a low pass system. However, it is clear that the PDI response has a higher bandwidth since the attenuation is significantly smaller at a given time corresponding to a specific frequency component for both experiments. The 3db point (bandwidth) was estimated by comparing the chirp response amplitude to the 0.1 Hz reference gain (Figure 4.3 and Figure 4.4).



Figure 4-4: 1st row: Vertical Response of the P and the PDI controllers to the chirp input signal. Upper left is the conjugate eye movement (black thin line) of the P controller to a 0.1 Hz sinusoid target (thick gray line). Upper right is the conjugate eye movement (black thin line) of the P controller to a 0.1 Hz sinusoid target (thick gray line). 2nd row: Input signal (laser dot). It is a sinusoid that increases its frequency from 0.1 Hz to 2 Hz within 90 sec. 3rd row: The conjugate Eye Response of the P controller. Black straight line: 3db point. 4th row: The conjugate Eye Response of the PDI controller. Black straight line: 3db point.

By estimating using linear interpolation the envelope of the conjugate eye response, a Bode plot for both controllers and both dimensions can be drawn. The two Bode plots are presented in Figure 4-5 and Figure 4-6. The increase of the bandwidth of the PDI controller compared with the one of the P controller is clearly observable.

The noise at high frequencies that is discernible in the Bode plots is due to the low sampling rate (30 Hz) and the phase lags introduced at high target frequencies. The gain envelope of the eye response is estimated by interpolating the maximum peaks of the sinusoidal eye trajectories in the chirp case. As the frequency of the chirp signal increases, both the amplitude and peak access become less accurate. (For a frequency of more than 1.6 Hz, less than 7 samples per period are being recorded.) What is more, phase lags at higher frequencies enhance the noise in gain estimate for frequencies above 1.5 Hz.



Figure 4-5: By estimating the envelope of horizontal conjugate eye response presented in Figure 4-3, the Bode plot of the P and the PDI controllers is created. The Bode plot for the P Controller (black line) shows a 3db point near 0.98 Hz while the Bode plot for the PDI controller (gray line) indicates a 3db point near 1.65 Hz.



Figure 4-6: By estimating the envelope of vertical conjugate eye response presented in Figure 4-4, the Bode plot of the P and the PDI controllers is created. The Bode plot for the P Controller (black line) shows a 3db point near 0.9 Hz while the Bode plot for the PDI controller (gray line) indicates a 3db point near 1.5 Hz.

The bandwidth of the P controller was estimated near 0.95 Hz and 0.9 Hz for horizontal and vertical dimensions respectively while the bandwidth of the PDI controller was estimated near 1.65 Hz and 1.55 Hz for the same dimensions. These numbers are in agreement with the simulations that have been presented in [60] with accuracy of less than 0.1 Hz.

4.6.1.2. Step Response

The step response of the slow phase mode (bilateral structure) for conjugate eye position is also illustrated in Figure 4-7. In order to obtain the step response, the eyes are simply driven from their initial location to a second position, by stepping the laser target. The step response, as well as the time constant, resemble those of the theoretical simulation in [60]. The observed time constant is approximately 0.240 seconds for the P controller and 0.110 seconds for the PDI controller. This value is close to the simulated time constants simulated in [60] and [91].

The faster rise time is achieved at the cost of a greater overshoot. However, it must be noted that in the complete system, step errors would be nulled by the saccadic (fast phase) mode and recruit slow phase only when the error is small. In these cases, the P and the PDI control trajectories are very similar (for t > 1.5 s in Figure 4-7). The main benefits of the PDI are in the improved gain and phase at higher frequencies.

4.6.1.3. Phase Lag (0.2 Hz and 1 Hz) and Retinal Errors

To test pursuit further, two different sinusoids of approximately 60 pixels (4 deg) amplitude are individually applied to the system. The results for horizontal movement of the target are illustrated in Figure 4-8 and Figure 4-9.

It is worth observing the attenuation of the response of the controller. In the P controller, the sinusoid output is significantly decreased at the higher frequency while the PDI controller is able to keep the energy of the signal from the input to its output. Another observable characteristic is the phase lag of the controller's output with respect



Figure 4-7: Step Response of P (black line) and PDI (gray line) Pursuit Controller. The rise time of the PDI controller is observed to be smaller than the rise time of the P controller.



Figure 4-8: Upper Panel: Pursuit Trajectory of P (left column) and PDI (right column) Controller following a slow target. Thick Gray line: Target. Black Thin line: Conjugate Eye movement. Bottom panel: conjugate error signals. The noise in the error signal is a result of the changing light conditions and centroid computations at sub-pixel accuracy (1 pixel = 0.05 deg)



Figure 4-9: Upper Panel: Pursuit Trajectory of P (left column) and PDI (right column) Controller following a rapid target. Thick Gray line: Target. Black Thin line: Conjugate Eye movement. Bottom panel: conjugate error signal.

to the input. In the P controller, the phase lag is significantly larger (this concords with the anticipated cutoff of approximately 1 Hz.) Due to the phase delay, the retinal error in the P controller is expected to be larger than that of the PDI. The retinal errors can be found in the second row of the following figures.

4.6.1.4. Two – dimensional Target

In order to test both horizontal and vertical controllers simultaneously, 2D targets are used as system inputs. The target consists of a circle while the input frequencies correspond to the angular frequency of the 2 D target. The test is performed within the operating bandwidth of both the controllers with a 0.2 Hz target in a circular form. The laser target executed a circular path on the wall (2D), in front of the robotic head.

In the following figures, the horizontal and vertical conjugate positions and errors are illustrated. An x-y plot is also provided showing the target and the conjugate eye position. The eyes clearly follow the 2 D target much better in the case of the PDI (Figure 4-12) controller in which the retinal slip is computed. The P controller response is provided in Figure 4-10.

To see how the pursuit is being executed with respect to time, Figure 4-11 and Figure 4-13 project the horizontal and vertical eye positions with respect to time as the third dimension. At this testing frequency, the retinal errors are in agreement with the gain of the predicted transfer functions for both controllers. Although the frequency of the target rotation is small, the attenuation of the P controller's response is observable. The retinal errors also in the case of the PDI controller are smaller than the ones calculated in the P controller.



Figure 4-10: 2 D target pursuit of P controller to 0.2 Hz input. On the left individual horizontal and vertical conjugate eye position (black) and error (gray) are illustrated; on the right the x-y plot is illustrated. The gray thick line is the laser target.



Figure 4-11: 2 D target P controller pursuit vs. time for 0.2 Hz input. The inner signal is the conjugate retinal error e_C (gray), and the outer signal s the conjugate position E_C (black).



Figure 4-12: 2 D target pursuit of PDI controller to 0.2 Hz input. On the left individual horizontal and vertical conjugate eye position (black) and error (gray) are illustrated; on the right the x-y plot is illustrated. The gray thick line is the laser target.



Figure 4-13: 2 D target PDI controller pursuit vs. time for 0.2 Hz input. The inner signal is the conjugate retinal error e_C (gray), and the outer signals represents the conjugate position E_C (black).

4.6.2 Switching Enabled

By enabling saccades, we can increase the controller's bandwidth to the point where it is even underdamped. Since the controller will alternate from saccade to pursuit mode once the estimated error is within a certain threshold, a very sharp transition can be observed, followed by smooth pursuit. Since the pursuit controller switches off before the error reaches the overshoot stage, the overall controller can have a much larger bandwidth and avoid overshoot in either mode. Unfortunately, given the 30 Hz sampling frequency, the current saccade mode is so fast, that the entire saccade only lasts a few (most of the times, a couple of) samples, making it difficult to return to pursuit before ringing has started in the saccadic mode. Hence, to obtain faster saccades without ringing the saccadic response is designed here not to overshoot.

4.6.2.1 Ramp Trajectory

To test the bilateral controller with fast phases enabled, its response on a ramp horizontal input is observed. The trajectories and the conjugate eye movement for the P and the PDI controllers are shown in Figure 4-14. The robotic head is fixating on a spot and suddenly the target makes a horizontal jump and follows a right direction with a constant speed.

As a result, after the initial saccade, the P controller does not take advantage of retinal slip (target speed). Hence, the conjugate line of sight lags behind the target and three catch-up saccades are required in the presented time interval (Figure 4-14).

In the trajectories of the PDI controller however, target speed prediction provides more accurate ramps and the cameras follow the target very accurately. No catch – up saccades are needed after the initial target jump since the pursuit mode is able to keep the retinal error within the tight limits of 1 degree deviation from the real target.

In the upper part of Figure 4-15 a trajectory of the controller response to a fast moving target is given: While a slow moving target (like the one in figure 4-14 that is moving with speed near to 1.7 deg/s) requires only the initial catch-up saccade, a fast moving target (like the one in figure 4-15 the speed of which is approximately 8 deg/s)

requires saccades periodically. As it is expected, the saccades in the case of the fast target trajectory are due to the saturation of the slip component, ie. the robotic eyes have some physical (and biological) limitations with respect to their moving speed.

The response of the controller to the ramp is approximated by the part of a sinusoidal input that has almost constant moving speed, as shown at the bottom of Figure 4-15. In the same figure, it can be observed that a reduction of the target's speed (in this case, during the peaks of the sinusoidal movement) results in a decrease in the frequency of the occurring saccades.



Ramp Trajectories for P and PDI Controllers (Low Speed Ramp)

Figure 4-14: Response of the P and PDI controller to a horizontal conjugate ramp input with low speed. Light Gray thick line: target trajectory. Black thin line: conjugate eye position. Gray thick line: Saccade. Saccade flag is also visible. The P controller after 11 s cannot keep up with the target and needs three saccades to stay close to the target in the given time period. The PDI controller is shifted 5s to the right for display purposes. After the first saccade, the PDI controller is able to track the target nicely and stay on it without need for any saccade.

4.6.2.2 Two – Dimension Examples

As originally shown for pure pursuit, results are illustrated here for horizontal and vertical motions, with fast phases enabled. Once again the target is an ellipsoid of 0.2 Hz. The experimental results for the P controller are provided versus time in Figure 4-16 and Figure 4-17. The equivalent plots for the PDI controller are shown in Figure 4-18 and Figure 4-19.



Figure 4-15: Upper Panel: The PDI controller's conjugate eye response to a fast moving ramp. The saturation of the slip leads to the introduction of periodical saccades in order for the cameras to catch up the target. Bottom Panel: The complete experiment showing the sinusoidal target trajectory and the conjugate response to it. The saccade flag indicates that as the target reaches the peak of its sinusoidal movement –and hence, its speed decreases- the controller is able to track the target effectively without any saccades (no slip saturation observed). The frequency of the saccades increases as the target speed (thus the retinal slip) increases. Light Gray thick line: target trajectory. Black thin line: conjugate eye position. Gray thick line: Saccade flag also observable.



Figure 4-16: 2 D target of P controller to 0.2 Hz input with saccades allowed. On the left, individual horizontal and vertical conjugate eye position (black) and error (gray) are illustrated; on the right the x-y plot is illustrated. The thick gray line is the laser target.



Figure 4-17: 2 D target P controller conjugate response vs. time for 0.2 Hz input with saccades enabled. The inner signal is the conjugate retinal error ec (gray), and the outer signal s the conjugate position E_C (black).



Figure 4-18: 2 D target of PDI controller to 0.2 Hz input with saccades allowed. On the left, individual horizontal and vertical conjugate eye position (black) and error (gray) are illustrated; on the right the x-y plot is illustrated. The thick gray line is the laser target.



Figure 4-19: 2 D target PDI controller conjugate response vs. time for 0.2 Hz input with saccades enabled. The inner signal is the conjugate retinal error ec (gray), and the outer signal s the conjugate position E_C (black).

The previous figures illustrate compound use of the horizontal and vertical modes of the controller. This means that if the vertical controller requires a saccade, the horizontal controller can remain in pursuit mode if the horizontal retinal error is below the ON threshold. Once again, the prediction that is present on the PDI controller is visible in Figure 4-18.

It is worth mentioning that the PDI controller can accurately track a moderate speed target even if the saccadic controller in one of the two dimensions is suspended. Experiments were made in which the horizontal controller could switch between fast and slow modes while the vertical controller was in pursuit mode exclusively. The resulted trajectory showed a really good tracking capability of the robot while the retinal error for both directions was kept in comparable levels with the errors seen in Figure 4-18.

An analogy of that experiment can be found in a clinical condition in which a lesion blocks the firing of some burst cells that drive the saccades. It is known from the literature that the pons cells' decisions on switching between the saccade and the pursuit mode do not take into account the direction of the target's movement. However, there are different burst neurons that drive the muscles during a saccadic movement for each one of the two directions. Hence, it is possible that the patient could have a descent tracking capability even if his bursts cells that drive the saccade in one dimension are being suppressed by a tumor.

4.6.3 Gaze

Recall that the gaze angle is simply the addition of the head angle to the conjugate eye angle as displayed in equation (4.11):

$$G=E_{c}+H \tag{4.11}$$

This section illustrates results when the neck is added to the binocular controller. This not only increases the system's bandwidth, but also models the human visual system more closely. In all the experiments that are presented, the saccades are enabled.



Figure 4-20: Upper Panel: Horizontal Gaze Response of the P controller. Gray thick line: target. Dashed line: Estimated Head position. Black line: Conjugate Eye movement. Black thick line: Gaze. Gray thick line: saccades. Bottom Panel: Conjugate Retinal Error.



Figure 4-21: Upper Panel: Horizontal Gaze Response of the PDI controller. Gray thick line: target. Dashed line: Estimated Head position. Black line: Conjugate Eye movement. Black thick line : Gaze. Gray thick line: saccades. Bottom Panel: Conjugate Retinal error.



Figure 4-22: Upper Panel: Horizontal Gaze Response of the PDI controller to a fast moving sinusoid target. Gray thick line: target. Dashed thin line: Estimated Head position. Black thin line: Conjugate Eye movement. Black thick line : Gaze. Gray thick line: Saccades. Gray thin line: Saccade Flag Bottom Panel: Conjugate Retinal error.

Movements will be presented in the horizontal plane. Comparable results have been observed for the vertical plane.

In Figure 4-20 and Figure 4-21 the gaze responses of the P and the PDI controllers respectively are shown. The prediction that is introduced with the PDI controller allows a much smoother overall tracking of the target. For both the experiments presented the gain of the head, K_{HEAD} is increased to the value 3, so that the

contribution of both the eyes and the neck to the gaze trajectory will be almost equal. Increment or decrement of the K_{HEAD} increases or decreases the amplitude of the sinusoid movement of the head respectively.

In Figure 4-22, the same experiment is shown using a much faster target (the speed as a function of the sinusoid frequency and amplitude is significantly increased). The high target speed results in a larger phase delay between the conjugate eye response and the target trajectory. Hence, the effect that the saccades have on the overall head-eye trajectory is obvious: The controller switching strategy automatically adapts to target speed and retinal error by dynamically changing saccade frequency and saccade duration.

The result is on average a reduced tracking error compared to what would be seen with a pure pursuit. The effect that the saccades have on the overall head-eye trajectory is obvious: The fast phase, exploiting the information of the retinal slip, drives the system's gaze to land perfectly on the target. The saccade flag indicates that as the target reaches the peak of its sinusoidal movement -and hence, its speed decreases- the controller is able to track the target effectively without any saccades (no slip saturation observed.)

That concludes the presentation of the experimental results that showed a significant improvement in the tracking capabilities of the controller. In the following chapter, the PDI controller will be capable of not only accurately following a target but also choose it among many distractors that will be present in its field of view.

Chapter 5 Towards Target Selection and Visual Attention

5.1 Introduction

In this chapter, a model for the bottom-up control of visual attention in primates is presented and experimentally tested in the robotic device previously described. Given an input image, the system attempts to predict which location in the image will automatically and unconsciously shift a person's attention towards it. In this biologicallyinspired system, an input image is decomposed into a set of "feature maps" which extract local spatial discontinuities in the modalities of intensity and velocity. Both feature maps are then combined into a unique scalar "saliency map" which encodes for the salience of a location in the scene irrespectively of the particular feature which detected this location as conspicuous. A winner-take-all system then detects the point of highest salience in the map at any given time, and draws the focus of attention towards this location.

In addition, the state of alertness of the robot is a function of the intensity perceived from the most salient target. The system's bandwidth is increased for brighter spots making the time required for the gaze shift smaller than the one needed to foveate a dimmer target.

5.2 Target Selection

Attention as a selective gating mechanism is often likened to a spotlight [75], enhancing visual processing in the attended ("illuminated") region of a few degrees of visual angle [87]. In a modification to the spotlight metaphor, the size of the attended region can be adjusted depending on the task, making attention similar to a zoom lens [92]. Neither of these theories considers the shape and extent of the attended object for determining the attended area. This may seem natural, since commonly attention is believed to act before objects are recognized. However, experimental evidence suggests

that attention can be tied to objects, object parts, or groups of objects [85]. How can we attend to objects before we recognize them?

We have developed a model (Figure 5-1) that estimates the extent of salient objects solely based on bottom-up information, serving as an initial step for subsequent object detection. Our attention system is based on the Itti et al. [45] implementation of the Koch and Ullman [52] saliency-based model of bottom-up attention. The model's usefulness as a front-end for object recognition is limited by the fact that its output is merely a pair of coordinates in the image corresponding to the most salient location. We introduce a method for extracting the image region that contains the attended objects from low-level features with negligible additional computational cost.

The model is developed in accordance with the known anatomy and physiology of the visual system of the macaque monkey as presented in Figure 2-7 and in Section 2.5. It comprises two interacting stages:

1. A fast and parallel pre-attentive extraction of visual features across the spatial maps (for brightness and motion). The brightness feature is computed by calculating the mean intensity of each blob detected inside the visual field of the robot. The motion is approximated by taking the first derivative of the X and Y Center of Gravity (CoG) coordinates for each blob.

2. A changing speed focal attention shifting mechanism, using a Winner-Take-All mechanism to select the most conspicuous image location.

The link between the two stages is a *Saliency Map*, which topographically encodes for the local conspicuity in the visual scene, and controls where the focus of attention is currently deployed. The saliency map directly receives its inputs from the preattentive, parallel feature extraction stage. Supervised learning can be introduced in a future work to bias the relative weights of the features in the construction of the saliency map and achieve some degree of specialization towards target detection tasks.



Figure 5-1: The Bottom-Up control of Visual Attention Model used for target selection.

The controller, with the addition of the multiple target detection feature, is able to handle a single target without any difference from the controller discussed in Chapter 4. However, when two or more targets are presented on the robot's visual field, a new process is introduced in sequence with the already existing computations. The output of that process is the Center of Gravity (COG) coordinates of the most salient target. Hence, the process of target tracking as described in Chapters 3 and 4 resumes as previously shown: the coordinates are given to the controller and the retinal errors are calculated and are fed into the controller's mapping functions. The process' flow-chart is presented in Figure 5-2.

When multiple targets are found, the selection of the most salient one is based on the calculation of the value $C_K = \alpha \Box Intensity_K + \beta \Box \left(\sqrt{Speed_X^2 + Speed_Y^2}\right)_K$ for each blob K.

Chapter 5



Figure 5-2: Flow- Chart of the Target-Selection Mechanism. While there is only one blob located, the controller is functioning exactly like it has been described in Chapter 4. When a second blob is visible, the center of gravities of the two blobs are being stored and a flag is set in order to detect in the next iteration that a second blob is indeed present and that it was not a misinterpretation of the frame grabber's blob analysis process. At the same time, the COG of the two blobs are being stored in order to be used for the calculation of the first derivative of the COG (an approximation of the target's speed). The blob that has the largest C_K is considered as the most salient one.

The parameters α and β can have different values depending on the feature (intensity or speed) that we want to favor. In the experiments to follow, the different spots' intensities have to have a small standard deviation (10 pixels) with a mean value of 245 in order to detect as small blobs as possible to avoid perturbations, under the presence of unavoidable recording noise. Hence, the speed was selected as the dominant saliency factor by using $\alpha = 0.1$ and $\beta = 0.9$.

5.3 Alertness of the Robot

For more than 100 years, lighting research has pursued the obvious - that with light we can see, without it we cannot. This pursuit has produced a dramatic increase in the understanding of how lighting affects vision, especially after the development of a special field of biology, named Photobiology. Photobiology studies the interaction of the light with the living organisms. Lighting enables people to see what they need to see, quickly, easily, and without discomfort. There are light installations that attempt to create specific impressions and, therefore, generate specific feelings in the viewer. Examples of these installations are found in churches, restaurants, shops and hotels, and outdoors as floodlit buildings or urban enhancement projects.

As explained in Chapter 2, light reaching the retina of the eye is converted into electrical signals that are transmitted by the optic nerve. Most of these signals end up in the visual cortex of the brain and produce our sense of vision. However, some of the nerve fibers split off from the optic nerve soon after leaving the eye and send signals to the suprachiasmatic nucleus (SCN), which is the area of the brain where the main clock for the human body resides. It is important to appreciate that this transfer of signals from the retina to the SCN is not a "branch line" of the visual system. It is independent and much older in terms of development than the visual system.

The signals that reach the SCN emanate equally from all parts of the retina without attempting to preserve their point of origin. While most of the visual cortex is devoted to the central two degrees of the retina and the projections from the retina are organized to preserve the retinal image, the nerve fibers and signals projecting onto the SCN do not attempt to preserve the image of the outside world reaching the retina. Hence, the circadian system is an example of a retinal-illumination detector.

In terms of physiology, the optic tract enters the diencephalon and extends to the lateral geniculate nucleus ("G" on the image pathway drawing beside). Some tract axons simply pass through the geniculate without synapsis on their way to motor centers (lateral arrows, as dictated from their color code, the motor projections correspond to perceived fields rather than being true to eye). This tract is a mixture of left eye/right eye axons; the geniculate is a 6-layered nucleus, and the fibers are sent to differing layers from each eye.



Figure 5-3: Image Pathway (modified by [117])

The layer/eye relationship is firmly maintained, so that each eye projects exclusively to three of the six layers on each side. The lateral geniculate merges the images acquired from the two eyes into one seamless picture before forwarding (post-synaptically) the visual data onward to the visual cortex. There is a visuotopic representation, "local visual sign" of the projection of ganglion cells to the geniculate, but the image is distorted in this representation by the demographics of retinal columns.

The retina projects to only about 10-20% of geniculate cells. Two layers of large cells are believed to be devoted to movement of the image in the two eyes. The remaining layers are of smaller neurons (cell bodies) and analyze the two images for color and for picture details. The remaining majority of input to the geniculate comes from other brain regions. This data apparently influences the projection to the visual cortex. Part of the afferent



Figure 5-4: Mapping of Mean Pixel Blob Intensity to Kverg (upper panel) and Kconj (bottom panel)

geniculate inflow is from the reticular system. Among other functions, this gigantic and diffuse mass of neurons governs the level of consciousness (and attention) as well as sleep (during which humans are functionally blind) and dreaming. At least a certain amount of this circuitry is a feedback loop. Hence, it can be argued that *not only does the level of alertness affect what we "see" but also what we see affects our level of alertness and concentration*.

In our biomimetic robotic implementation, this fluctuating level of alertness can be implemented by altering the gains of the proportional (P) part of the PDI controller. As it has been experimentally proven and shown (Appendix B), the increase of the K_{conj} and K_{verg} affects the overall system's bandwidth, hence its state of alertness since a system with a narrower bandwidth responds more sluggishly to a step input, ie. a non-expecting target that is suddenly presented on its visual field. In the experiments to follow, the K_{conj} and K_{verg} are functions of the *perceived intensity* which is defined for the purposes of this thesis as the *mean value* of the intensities of the pixels for each blob. As it can be observed in Figure 5-4, a linear relationship between the mean intensity of the salient blob and the K_{conj} and K_{verg} has been chosen. That being said, the system's bandwidth is in the range of [0.4, 2.6] Hz.

5.4 Experimental Results

The experiments to follow are used to test the presented theoretical analysis and prove that the robot is able for the first time to decide upon which target to follow based on the speed and the brightness of each visible target. The controller's parameter set used for this chapter's experiments can be found in Appendix B.

5.4.1 Two Targets with Different Velocities and the same Brightness

In the experiment presented in Figure 5-5, two targets of approximately the same mean intensity are inserted into the robot's visual field. One target is moving in a sinusoidal horizontal trajectory and the other is being kept stable in a level that is lower by nearly 2 degrees than the first one.

The first (moving) target is not present throughout the entire experiment. The function generator that controls the target's vertical position on the wall is being shut off for a certain time period, driving the laser dot to a spot that is not visible to the robotic eyes. For these periods, only one (the stable) target is visible and the eyes fixate on it. As soon as the moving target is re-introduced to the visual field, the eyes leave the stable spot and follow the fastest (and hence the most salient) target.

The same experiment in 2-D is shown in Figure 5-7. The position of the second (unmovable) target relative to the position of the moving target as well as the conjugate eye response trajectory can be seen.



Figure 5-5 : 3-D plot of the experimental response of the horizontal pursuit controller under the presence of two targets. The controller follows the fast moving target and only on the absence of that target it fixates on the stable target. As soon as the moving target resumes, the controller returns back on it, leaving the stable (and hence less salient) target unattended. Light gray Line: 1st visible target with Sinusoidal movement. Black Line: 2nd visible target that is kept stable during the time of the experiment. Dark gray Line: Conjugate Eye Response Trajectory.

5.4.2 Effect of the Brightness of the Target on the Step Response of the Pursuit Controller

As explained in Section 5.3, the alertness of the robot (ie its response speed) can be a function of the target's features (see Figure 5-4). In Figure 5-8, various step responses of the horizontal pursuit controller are displayed. For each step response, different intensity targets have been used. It is clear that the brightest target can produce the fastest horizontal conjugate eye response compared with the other step responses. Comparable results have been found for the vertical dimension.



Figure 5-6 : Individual Horizontal and Vertical Dimension of the experiment shown in Figure 5-5. Dark gray Line: 1st visible target with Sinusoidal movement. Gray dotted Line: 2nd visible target that is kept stable during the time of the experiment. Black Line: Conjugate Eye Response Trajectory.



Figure 5-7 : 2-D plot of the experiment shown in Figure 5-5. The 2-D position of the second (stable) target has been indicated with a circle.



Step Response of Horizontal Pursuit Controller for Step Input with Different Intensities

Figure 5-8: Step Response of the Pursuit Horizontal Controller for Various Intensities Step Inputs. Light thick Line: Target trajectory. Black Line: Conjugate Eye Response trajectory for a target of intensity equal to 255. Dark gray Line: Conjugate Eye Response trajectory for a target of mean Intensity equal to 250. Lighter gray Line: Conjugate Eye Response trajectory for a target of mean Intensity equal to 245. Lightest gray Line: Conjugate Eye Response trajectory for a target of mean Intensity equal to 245. Lightest gray Line: Conjugate Eye Response trajectory for a target of mean Intensity equal to 245. Lightest gray Line: Conjugate Eye Response trajectory for a target of mean Intensity equal to 240.

5.4.3 Controller Response under the presence of two targets with different intensities and velocities

In the last experiment, displayed in Figure 5-9, a very bright (having mean intensity equal to the maximum 255) flash light spot is used as a stable target and a laser dot is used as a moving target with mean intensity equal to 250. The unmovable spot is fixed in the plane of the second target's horizontal movement and is located nearly 5 degrees away from the right sinusoid peak.

It is easy to understand that as the target is moving near the peak of its sinusoid trajectory, its speed decreases until temporarily nulled (at the local maximum of the movement). Then the target (and as it starts its descent towards the origin) speeds up.



Figure 5-9: Experimental Response of the Horizontal Pursuit Controller under the presence of two targets with different intensities and velocities. The eye conjugate trajectory follows the fast moving target until the moment that the combination of the intensity and speed as defined by the C_1 value in Figure 5-2 is no longer larger than the equivalent C_2 of the 2nd target. Gray dashed Line: 1st visible target with sinusoidal movement. Dark gray Line: 2nd visible target that is brighter than the 1st target and which is kept stable during the time of the experiment. Black Line: Conjugate Eye Response Trajectory.

The controller follows the laser dot accurately during the movement of the target in the left plane with respect to the origin. However, and as the conjugate eye trajectory approaches the right peak of the sinusoid (and its speed decreases gradually to zero), the value $C_K = \alpha \Box Intensity_K + \beta \Box \left(\sqrt{Speed_X^2 + Speed_Y^2}\right)_K$ for the moving target becomes smaller than the corresponding C_K value for the stable but brighter flash light dot. Hence, the pursuit controller response is driven further from the peak of the sinusoid and lands on the unmovable target.

The eyes stay on the stationary dot as long as the speed of the moving target makes the C_k value smaller than the one calculated for the stable target. As the sinusoid target moves away from the peak of its trajectory, it speeds up and when the speed reaches a certain threshold, the controller decides that the moving target has become

again the most salient target. The changes on the controller's decisions regarding the most salient target can be observed by the sudden peaks introduced in the conjugate retinal error at the moment when the eyes leave one target for the pursuit of the other.

5.5 Conclusion

In this chapter, the biomimetic nature of the robot was further enhanced by allowing for the first time the parallel observation of two targets and the selection of the most salient one with respect to its intensity and speed. In all circumstances, the results affirm that the controller is able to judge about the significance of a target that suddenly comes into its visual field. The next and final chapter will conclude this thesis by summarizing the results in more detail and proposing future work.

Chapter 6 Conclusions

6.1 Synopsis and Discussion

The goal of this thesis was to enhance the robotic gaze controller implemented by Sejean [91] by adding prediction and memory and by using the speed and the brightness of the target as a significance factor in order to distinguish it from various distractors. These improvements, having been theoretically justified, experimentally tested and concisely presented, can now be recapitulated:

The objective to increase the time that the controller operates in slow phase mode was met by the introduction of the PDI control. Physiological findings of visual pathways that use prognosis and memory for the accurate estimation of the target's position guided the design of mapping functions inside the sensory-motor box to include the computation of the first derivative of the retinal error superimposed on an integration over time version. The experimental results of the PDI controller showed that the hardware device is now capable of fixating a target significantly better and over a larger bandwidth compared with the P controller implemented by Sejean [91]. The robotic eyes can detect and track a faster moving target with a small number of randomly occurring saccades. Similar performance even with random inputs, subject to the filtering properties provided in results (Fig. 4-5 and Fig. 4-6.)

Since retinal slip information is now available to the controller, more accurate programming of saccade amplitude programming could also be achieved. Published results provided the neurological base on which the inclusion of the retinal slip lies. The rate of change of slip is also monitored to prevent the controller from extreme overshoots at points of discontinuities or high curvatures in target trajectories, during both slow pursuit and saccades. Concerning the switch between the two modes of the controller, the decision for the transition from pursuit to saccade mode is now altered so that the information of the retinal slip is also embodied: A faster target would favor the transition from pursuit to saccade mode while a slow moving one would make that transition less likely to happen. In addition, a velocity threshold is introduced to limit the saccade duration in cases where the controller's response amplitude reaches a plateau while not having yet accomplished its saccade goal.

The parameters presented in Lee's theoretical work [60] provided the controller with a larger but non-ideal gain profile at higher frequencies. A new analysis of the controller's parameters results in an alternative parameter set that provided an even larger and also flatter bandwidth. Furthermore, a target can now be tracked in a scene that has multiple distractors: The first derivative as well as the intensity of each blob detected inside the visual field of the cameras are calculated and the most significant blob with respect to its speed and brightness is considered as the dominant target, the one to which the robotic device turns its attention. The attention is also altered with respect to the intensity of the target so that a brighter blob would result in faster response of the controller while a dimmer one would slow down the alertness of the robot.

This last approach relies on the known fact that a brighter environment raises alertness levels in humans and that alertness is associated with faster responses (larger bandwidth), smaller delays and more accurate step responses – presumably all allying on simple recruitment levels (greater sensitivity to sensory drives with alertness.)

6.2 Significance

The significance of the above results can be viewed in terms of the possible applications that the controller is now able to have. The field of surgery is entering a time of great change, spurred on by remarkable recent advances in surgical and computer technology. Robotic devices could be used in more than 3.5 million medical procedures per year in the United States alone. The first generation of surgical robots is already being installed in a number of operating rooms around the world since 1994: The Da Vinci

Surgical System [115], the ZEUS Robotic Surgical System for minimally invasive microsurgery procedures and the voice-controlled Automatic Endoscopic System for Optional Positioning (AESOP) [114] are three surgical automated tools that have been recently developed while NeuroArm ([63], [112]) is the first robot created for the specific purpose of performing micro-neurosurgery. While surgeons are trained to work at an accuracy level of about 1 millimeter, such robots can be accurate to 30 microns. The controller presented in this thesis could become the heart of such a robotic surgeon introducing a better fixation of the cameras to the various "targets" of an operation. This increased accuracy can prove particularly important today and, together with the advances in nanotechnology and robotics, will afford the ability to shift surgery from the organic to the cellular level.

If the doctor doesn't have to stand over the patient to perform the surgery, and can remotely control the robotic arms at a computer station a few feet from the patient, the next step would be performing surgery from locations that are even farther away. The use of a computer console to perform operations from a distance opens up the idea of *telesurgery*, where a doctor performs delicate surgery miles away from the patient. A major obstacle in tele-surgery has been the time delay between the doctor moving his or her hands to the robotic arms responding to those movements. (Currently, the doctor must be in the room with the patient for robotic systems to react instantly to the doctor's hand movements.) The same delay problem can be involved in any application that would require remote control – mars exploration rovers, bomb defusion robotic devices etc. Our controller, being able to track multiple targets and separating one from its distractors, could now guide such a robotic device using saccades and pursuit movements without the need of being controlled for each one of the movements. Hence, the control of the position of the device can now be shifted from the hands of its operator to the robotic device itself, when a preferred spatial goal is determined.

While surgical robots offer some advantages over the human hand, we are still a long way from the day when autonomous robots will operate on people without human interaction. But, with advances in computer power and artificial intelligence, it could be that in this century a robot will be designed that can locate abnormalities in the human body, analyze them and operate to correct those abnormalities without any human guidance.

Another application in which the presented controller can be found useful is that of face tracking. Security, being the primary goal of every governmental and private institution nowadays, urges for better identification techniques for safe access. One of these techniques is face recognition. In order to have a better recognition though, the image viewed from an individual's face, needs to be stable. The current controller, being able to track fast moving targets with no or limited number of saccades could support a more reliable authentication process, with more robust computations.

6.3 Future Extensions

Care was taken throughout this thesis to lay a solid foundation upon which further study of this and related fields could be based. In the last part of this final chapter, proposals for future research work are presented. Given the task-specific nature of many imaging problems as well as the thorough annotation of the different parts and devices of the robot, a clear understanding of the basic principles enhances significantly the chances for the successful solution of any future research problem.

6.3.1 Object Recognition

Starting with this thesis, the robot's treatment of digital image processing began a transition from processes whose outputs are images to processes whose outputs are attributes about images. The ultimate goal of any future work in the robotic device should be the use of the processing capabilities of the computer to emulate human vision, including learning and being able to make inferences and take actions based on visual inputs. Any future work towards that direction should aim for higher level processing that involves "making sense" of an ensemble of recognized objects and at the far end of the continuum, performing the cognitive functions normally associated with vision.

One of the most interesting aspects of the world is that it can be considered to be made up of patterns. Hence, an implementation of techniques for object recognition in the hardware device could guide the robot to survive in an unknown environment. At this point, the binocular robot reacts to the visual scene simply on the basis of brightness and of moving centers of gravity. Two principal approaches - decision-theoretic and structural – could be used in order for the robot to "learn" from sample patterns. The first category deals with patterns described using quantitative descriptors, such as length, area and texture. The second category deals with patterns best described by qualitative relational descriptors such as scene labeling, string grammars and similarity measures. The field of computer vision has many alternatives that could be merged with our innovative platform controller.

6.3.2 Higher Operating Frequency

In order for the visual processing and analysis tasks to be completed on time so that little control delay is introduced, a controller frequency should have a higher frequency than that of the image acquisition process. Increasing the controller's operation frequency has physiological justification: the internal brainstem circuits controlling eye movements operate with neural delays as small as 2ms whereas the visual processing pathways can have delays as large as 100ms – 200ms.

The higher operation frequency can be useful not only in terms of the (pseudo) parallel image processing load that the controller can handle, but can also increase the bandwidth of the saccadic system that is only restricted by the sampling interval since saccades are executed blindly with no true visual feedback. For this thesis, the controller frequency is equal to the frame grabbing frequency (30 samples per second). A faster working frequency can provide the controller with a better time resolution during the fast phase mode and permit switching strategies that avoid target overshoot. Similarly, larger bandwidths will be available to the saccadic mode without introducing underdamped responses.
6.3.3 Introducing 3D- Torsion Movement used for Targets with Dimensions

The cameras in this thesis can make two dimensional movements, namely yaw and pitch. An obvious extension of the current work would be to provide the cameras with a third actuator giving them the capability of 3D movement with the inclusion of torsion. The torsional movement of the robotic eyes will allow them to track accurately a target that does not have symmetrical dimensions (ie. it is not a circular blob like the targets that are used so far for the robot).

A consequence of such an improvement to the robotic device, would be the near perfect fixation of a target with random dimensions: The rotation of the two cameras around an axis that is perpendicular to their frontal plane (retina) would allow for better image fusion when the cameras are deviated from the null direction.

6.3.4 Hand – Eye Coordination

Hand-eye coordination is the ability of the vision system to coordinate the information received through the eyes to control, guide, and direct the hands in the accomplishment of a given task. It is well established that vision dominates all other sensory modalities in the control of movement and in motor learning. Motor control (even the most sophisticated one that undertakes delicate missions like operating a patient) is then usually based on the pre-computation of the different components of the system using inverse kinematics and the visual information.

However, by following the direction towards which the primate brain evolved over the ages, modifications can be made to the controller to relate visual signals to coordinated orienting hand and arm movements. In humans, motor coordination can be considered to take place in the Superior Colliculus (SC) which is an important relay center for different types of sensory-motor information important for orientation tasks.

The biomimetic logic can thus be extended to hand - arm control in order to research upon the information that is provided by eye movements and the means that this information is used for the guidance of hand movements. Adding end-effect control to any robotic platform (ie. allowing the controller to compute the required movement based only on the current goal error without having to compute each of its moving components separately) could boost up the accuracy and the speed provided by robots and provide safety measures (goal boundaries) to satisfy Asimov's first law of robotics [6].

Appendix A Nonlinear Calibration of the Laser Dot and the Cameras

The beam produced by the He-Ne Laser Projector is reflected by two mirrors on its way to a wall in front of the binocular robotic platform (Figure A-1). The first mirror is used to center the beam on the x-axis while the second one is used to move the laser dot vertically on the wall. In other words, the dot's position is controlled by two motors, one for each dimension. The two motors are driven by a voltage source which is selected to be either an external function generator or the 6052E analog output of the PC. Two voltage sources are needed.

With this experimental setup, the voltage that drives the laser beam onto the wall needs to be mapped to the 2-D angular position of the projected dot, with respect to the midline between the two cameras. The position of the wall that is at the same height with the cameras and in the middle of their horizontal distance is regarded as *the center of the axes* (ie the position in which the angle is taken as zero).

Furthermore, the way that the beam is projected onto the wall, as shown in Figure A-2 makes a linear calibration useless because the distance between the wall and the 2nd mirror gets bigger as the dot is moving away from the center of the axes. It can be observed that the angle at which the mirror is swiveled equals the angle at which the beam projects onto the wall. It follows that there is a trigonometric relationship between the voltage that is fed into the mirror rotating motors and the angle of the beam projected to the wall. By measuring the x-y coordinates of a spot on the wall and by calculating the corresponding angles using the equation $\Delta A = \tan^{-1}\left(\left(\frac{D}{d}\right) \Box \frac{180}{\pi}\right)$, where D is the distance from the origin($\Delta X, \Delta Y$) and d = 82 cm is the distance between the wall and the camera frontal plane, the corresponding angle can be found.



Figure A - 1 : Laser Beam Scattering. The laser is projected onto the wall with the help of two mirrors, the last of which is used to control the exact location.

However, the mapping function can be approximated, in the range of interest, as a 3^{rd} order polynomial of the form $\Delta A_x = \alpha_x (\Delta X)^3 + \beta_x (\Delta X)^2 + \gamma_x (\Delta X) + \delta_x$ for the horizontal dimension and as another polynomial of the same form $\Delta A_y = \alpha_y (\Delta Y)^3 + \beta_y (\Delta Y)^2 + \gamma_y (\Delta Y) + \delta_y$ for the vertical, where ΔX is the distance (in pixels or cm) between the current blob position and the origin in the horizontal axis, ΔY is the distance in the vertical axis and $\Delta A_{\chi,y}$ is the angle (in degrees) that corresponds to the current position.



Figure A - 2: The laser (gray line) is reflected by the mirror. The angle of the rotated mirror is equal to the angle of the beam that is projected on the wall.

The parameters α , β , γ and δ of the mapping functions can be easily computed by solving a 4 x 4 linear system that would use as ΔX (ΔY) distances that have been physically measured (using a measuring tape) and computing the angles using the aforementioned trigonometric equation.

The stand-alone C-code by National Instruments Labs, used to drive the laser dot from the Analog-to-Digital Output card, introduces an attenuation factor of approximately 0.5 to the pc driven voltages with respect to the voltage used from the function generator. Hence, the input mapping functions depend on whether the motors of the mirrors are driven from the pc output analog card or are controlled by external function generators. Also, a different mapping function is found for each dimension. The horizontal and vertical angles used to map the distance in pixels corresponds to the conjugate angle of the two cameras. The mapping functions are slightly asymmetrical and that can be explained by the geometrical position of the mirrors on a table surface *below* the head and the cameras.

Horizontal Dimension					
Point X axis	Volts		Left Camera	Right Camera	Conjugate Angle
	Function Generator	РС	ΔX (pixels)	ΔX (pixels)	
X ₁ 0 cm	0.042 V	0.080 V	0	0	0.00°
X ₂ +1.15 cm	0.083 V	0.180 V	-12	+14	0.80°
X ₃ + 2.30 cm	0.132 V	0.280 V	-23	+24	1.61°
X ₄ + 4.70 cm	0.233 V	0.480 V	-45	+47	3.28°
X ₅ + 6cm	0.288 V	0.580 V	-57	+60	4.19°
X ₆ +14.8 cm	0.57 V	1.280 V	-133	+139	10.23°
X ₇ +23.2 cm	0.94 V	2.000 V	-211	+224	15.79°
X ₈ +43.25cm	1.69 V	3.500 V	-368	+396	27.81°
X ₉ -3.2 cm	-100 mV	-0.180 V	+28	-28	-2.24°
X ₁₀ -7.9 cm	-292 mV	-0.580 V	+73	-71	-5.50°

Table A - 1 : Horizontal Dimension Calibration Measurements

The beam is projected onto the wall using mirrors the surface of which is approximately 1 cm^2 . Hence, a small physical displacement of the mirror with respect to its ideal (centralized) position has a consequence to the symmetry of the relationship between the voltage that moves the mirror and the position of the laser dot on the wall.

Vertical Dimension					
Point	Voltage Source		Left Camera	Right Camera	Conjugate
Y axis	Function	PC	ΔΥ	ΔΥ	Angle
	Generator		(pixels)	(pixels)	
Y ₁	1.910 V	3.950 V	0	0	0.00°
0 cm					
Y ₂	1.960 V	4.050 V	+20	+22	1.26°
+1.80 cm					
Y ₃	2.010 V	4.150 V	+38	+41	2.54°
+3.65 cm					
Y ₄	2.110 V	4.350 V	+74	+78	5.40°
+7.75 cm					
Y ₅	1.550 V	3.250 V	-108	-109	-8.25°
-11.9 cm					
Y ₆	1.360 V	2.850 V	-174	-171	-12.65°
-18.4 cm					

Table A - 2: Vertical Dimension Calibration Measurements

The mapping functions' parameters that are found with this process are shown in Table 3, their plots in the working range are shown in Figure A-3. In the first row of the figure, the mapping function for the conjugate eye position is presented. The controller run under the presence of a single target for a few seconds until the retinal error became practically zero (both robotic eyes landed on target). A sudden displacement of the laser dot caused by an instantaneous predefined increase of the voltage source in the horizontal (vertical) dimension introduced to the controller a conjugate retinal error of a certain amount of pixels. The retinal errors values, found in the 4th and 5th row of Table A-1 and Table A-2, provided the required information to calculate the mapping functions for the conjugate eye movement.



Figure A - 3 : Mapping Functions for Horizontal (left column) and Vertical (right column) Pixel distance (first row) and laser dot position (second and third row)

For the second (third) row, the same method of the laser step displacement is used. We measure physically (using a measuring tape on the wall) the distance between the two dots as well as the voltage coming out of the function generator (PC analog output card) to drive the mirror in two dimensions.

A simulated experiment showing better the non-linear mapping effects of the Function Generator Vertical Mapping function is presented in Figure A-4. In Table A -3, the α , β , γ and δ parameters for the 3rd order polynomial mapping functions are provided.



Figure A - 4 : Simulated Sinusoid Input Voltage received from the Function Generator (upper panel) and the output of the mapping function (bottom panel) in angles. The x-axis is discrete time with sampling frequency equal to 30Hz.

The mapping process that has been described is used only for display purposes. The controller is still running calculating the retinal error in terms of pixels. For circumstances in which the laser dot did not move far away from the origin, the nonlinear functions were considered redundant and linear mapping functions were used instead.

The mapping functions presented in the first row of Figure A-5, are used for both the conjugate eye position as well as the conjugate retinal error. In Figure A–5, an example is given for the mapping of a binary noise retinal error (shown in Figure 4-8) from pixels to degrees.

	Horizontal Dimension			Vertical Dimension		
	Function	PC	Conjugate	Function	PC	Conjugate
	Generator		Angle	Generator		Angle
α	1.1347	0.9940	0.0000	21.7094	3.6237	0.0000
β	-0.4814	-0.1010	-0.0002	-111.9932	-40.4779	0.0001
γ	16.5941	7.9862	0.0825	213.5795	161.9647	0.0630
δ	-0.5746	-0.6400	-0.2447	-150.5846	-231.4835	-0.1002

Table A - 3 : Parameters of the Mapping 3rd degree Polynomial Functions



Figure A - 5 : Sample Error in Pixels (upper panel) and in degrees (bottom level)

Appendix B Finding Optimal Parameters for PDI Controller

An analytical approach for finding the suitable parameters for the PDI model is impractical because of the complexity of its transfer functions. Therefore the parameters are found empirically. A thorough analysis of the effects of the various controller parameters on the flatness and the width of the system's bandwidth follows. The analysis will lead to the presentation of a better parameter set than the one proposed by Lee [60]. The parameter values found by Lee are used as a starting point. They are used as the centre points around which a search on a finer scale is conducted, for the optimal combinations which yield good frequency behaviors. In order to quantize the "quality" of a parameter set, the 3dB bandwidth, the transient overshoot in step response and the absolute % overshoot (or undershoot) of a sinusoidal test in the range [0 3] Hz are used. The following figures illustrate these empirically found relationships. (The horizontal vergence is used as an example)

In Figure B-1, Figure B-2 and Figure B-3, the empirically found relationships are illustrated. In order to be consistent with Lee's work, the horizontal vergence controller with average eye plants was used. The results can be summarized as follows:

- 1. When K_{verg} and K_{verg}^{slip} are fixed, the transient overshoot increases linearly with K_{verg}^{int} . The 3dB bandwidth increases linearly as well, but by very small amounts.
- 2. When K_{verg}^{slip} and K_{verg}^{int} are kept constant, the 3dB bandwidth increases linearly with K_{verg} .
- 3. When K_{verg} and K_{verg}^{int} are constant, the 3dB bandwidth decreases linearly as K_{verg}^{slip} increases.



Figure B-1 : Horizontal Vergence PDI Controller Bandwidth (upper panel), Step Response (left bottom panel) and Peak-to-Peak absolute over(under)shoot (right bottom panel) with respect to the proportional gain (P). The bandwidth increases as the P increases. After the P reaches a certain value, a small "bump" is introduced into the system's bandwidth. The step response shows a smaller time response as the bandwidth increases, yet with a larger overshoot. The Peak-to-Peak over(under)shoot shows that as the proportional gain increases, the absolute % becomes smaller until a certain value after which there is a large undershoot after 1.3 Hz.

Having observed the relationships between the PDI parameters and various parameters used to gauge the quality of the controller transfer functions, the task of finding optimal $(K_{conj}, K_{conj}^{slip}, K_{conj}^{int})$ and $(K_{verg}, K_{verg}^{slip}, K_{verg}^{int})$ sets become more simplified. The resulted parameter set is shown in Table B-1.



Figure B-2: Horizontal Vergence PDI Controller Bandwidth (upper panel), Step Response (left bottom panel) and Peak-to-Peak absolute over(under)shoot (right bottom panel) with respect to the slip gain (D). The bandwidth increases as the D increases. After D reaches a certain value, a small "bump" is introduced into the system's bandwidth. The step response shows a smaller time response as the bandwidth increases, yet with a larger overshoot. The Peak-to-Peak over(under)shoot shows that as the slip gain increases, the absolute % decreases until a certain value after which the % over(under)shoot increases.

The frequency responses, found analytically by using the parameter set presented in table B-1, are shown in Figure B-4 to Figure B-9. These figures also offer a comparison demonstrating the marked improvement in the bode diagrams from Lee's proposed parameter set.



Figure B-3: Horizontal Vergence PDI Controller Bandwidth (upper panel), Step Response (left bottom panel) and Peak-to-Peak absolute over(under)shoot (right bottom panel) with respect to the integral gain (I). The bandwidth is flatten as the I decreases. The bandwidth is no more flattened after 10⁻⁷. The step response shows an almost identical time response as the I decreases, yet with a smaller overshoot (that saturates also after 10⁻⁷). The Peak-to-Peak over(under)shoot shows that as the integration gain increases, the absolute % decreases.

Conjugate			Vergence		
	Horizontal	Vertical	Horizontal	Vertical	
K	4.53e-5	4.1e-5	3.9e-5	4.08e-5	
K _{slip}	-0.5e-8	-0.1e-8	-0.445e-6	-0.1e-7	
Kint	0.1e-6	1.3e-6	0.4e-5	0.9e-5	
g	0.001	-0.003	0.001	0.002	

Table B - 1 : New Parameter Set for Conjugate and Vergence Horizontal and Vertical Controller



Figure B-4: Horizontal Vergence Controller Bode plot for old parameter set (dashed line) and new parameter set (solid line). The bandwidth has been increased by almost 0.5 Hz and has become flat. Left column is for horizontal plane and right column is for vertical plane.



Horizontal Vergence Controller Step Response Vertical Vergence Controller Step Response Rise Time =[0.0003] - Overshoot =0.053683 Rise Time =[0.0003] - Overshoot =0.060786

Figure B-5 : Horizontal Vergence Controller Step Response for old parameter set (dashed line) and new parameter set (solid line). The rise time slot has been narrower by 0.2ms and the overshoot has become smaller too. Left column is for horizontal plane and right column is for vertical plane.



Figure B-6: Horizontal Vergence Controller Percentage of Peak-to-Peak over(under)shoot for old parameter set (dashed line) and new parameter set (solid line). The overshoot is much smaller in the new parameter set while the undershoot after 1.3 Hz is negligible. Left column is for horizontal plane and right column is for vertical plane.



Figure B-7: Horizontal Conjugate Controller Bode plot for old parameter set (dashed line) and new parameter set (solid line). The bandwidth has been increased by almost 0.2 Hz and has become much flatter. Left column is for horizontal plane and right column is for vertical plane.



Figure B-8: Horizontal Conjugate Controller Step Response for old parameter set (dashed line) and new parameter set (solid line). The rise time slot has been narrower by 0.1ms and the overshoot has become much smaller. Left column is for horizontal plane and right column is for vertical plane.



Figure B-9: Horizontal Conjugate Controller Percentage of Peak-to-Peak over(under)shoot for old parameter set (dashed line) and new parameter set (solid line). The overshoot in the new parameter set is negligible. Left column is for horizontal plane and right column is for vertical plane.

References

- Ahmed S., VISIT: An Efficient Computational Model of Human Visual Attention. *PhD thesis*, University of Illinois at Urbana-Champaign, 1991.
- [2]. Allman J., Miezin F., and McGuinness E. "Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area." *Perception*, 14:105--126, 1985.
- [3]. Aloimonos Y., et.al. "Action vision.", Int'l. J. Comp. Vision, 7:333-356, 1988.
- [4]. Arsenio, A., and Santos-Victor, J. "Active monocular tracking with temporal integration of visual cues", 9th Portuguese Conference on Pattern Recognition RECPAD'97, Coimbra, Portugal, 1997.
- [5]. Argyros, A.A., Bergholm, F. "Combining Central and peripheral vision for reactive robot navigation", *Computer Vision and Pattern Recognition Conference (CVPR)*, Colorado, USA, 1999.
- [6]. Asimov, Isaac. The Caves of Steel, Doubleday, 1952.
- [7]. Ballard D. "Animate vision.", Artificial Intelligence, 48:57-86, 1991.
- [8]. Barmack, N. H. "Modification of eye movements by instantaneous changes in the velocity of visual targets." *Vision Res.* 10:1431 - 41, 1970.
- [9]. Barreto, J.P., Peixoto, P., Batista, J., and Araujo, H, "Performance characterization of visual behaviors in an active vision system", SIRS'98-6th Symposium on Intelligent Robotics Systems, Edinburgh, Scotland, pp. 309-318, 1998.
- [10]. Basso MA, Krauzlis RJ and Wurtz RH. "Activation and inactivation of rostral superior colliculus neurons during smooth-pursuit eye movements in monkeys.", J Neurophysiol 84: 892–908, 2000.
- [11]. Batista, J., Peixoto, P., and Araujo, H. "Real-time vergence and binocular gaze control", *IEEE Proc. IROS*, pp. 1348-1354, 1997.
- [12]. Becker, W., Klein, H. "Accuracy of saccadic eye movements and maintenance of eccentric eye positions in the dark.", *Vision Res.* 13:1021-34, 1973.
- [13]. Bergeron A and Guitton D. Fixation neurons in the superior colliculus encode distance between current and desired gaze positions. *Nat Neurosci* 3: 932 939, 2000.

- [14]. Bergeron A and Guitton D. In multiple-step gaze shifts: omnipause (OPNs) and collicular fixation neurons encode gaze position error; OPNs gate saccades. J Neurophysiol 88: 1726–1742, 2002.
- [15]. Bergeron A, Matsuo S, and Guitton D. Superior colliculus encodes distance to target, not saccade amplitude, in multi-step gaze shifts. *Nat Neurosci* 6: 404–413, 2003.
- [16]. Bridgeman B. "Conscious vs. Unconscious processes: The case of vision.", Theory and Psychology, 2(1), 73-88, 1992.
- [17]. Brouwer de S., M. Missal, G. Barnes, and P. Lefevre, "Quantitative analysis of catch-up saccades during sustained pursuit." J. Neurophysiol., 97: 1772-1780, 2001.
- [18]. Brouwer de S, Yuksel D, Blohm G, Missal M, Lefèvre P. "What triggers catch-up saccades during visual tracking?" J Neurophysiol, 87, 1646-50, 2002.
- [19]. Brown, C. "Gaze controls with interactions and delays", *IEEE Transactions on systems, man, and cybernetics*, vol. 20, no. 1, pp 518-527, 1990.
- [20]. Carpenter, R. H. S. "Cerebellectomy and the transfer function of the vestibuloocular reflex in the decerebrate decerebrate cat.", *Proc. R. Soc. London Set. B.* 181:353-74, 1972.
- [21]. Carpenter, P. A. & Just M. A. "Reading Comprehension as Eyes See It.", Cognitive Processes in Comprehension, Carpenter & Just, Lawrence Erlabaum Ass., Hillsdale, 1977.
- [22]. Carpenter G.A., et.al. "The what-and-where filter: a spatial mapping neural network for object recognition and image understanding.", *Computer vision and Image Understanding*, 69(1): 1-22, 1998.
- [23]. Clark, J.J. and Ferrier, N.J. "Modal control of an attentive vision system", IEEE Proc. of the 2nd International Conference on Computer Vision, Tampa, FL, pp. 514-531, 1988.
- [24]. Cohen, B., Komatsuzaki, A. "Eye movements induced by stimulation of the pontine reticular formation: evidence for integration in oculomotor pathways.", *Exp. Neurol.* 36:101-117, 1972.

- [25]. Cohen, B., Matsuo, Y., Raphan, T. "Quantitative analysis of the velocity characteristics of optokinetic nystagmus and optokinetic after-nystagmus.", J. *Physiol. London.* 270:321-44, 1977.
- [26]. Collewijn, H. "Optokinetic eye movements in the rabbit: input output relations.", *Vision Res.* 9:117-32, 1969.
- [27]. Coombs, D.J., and Brown, C.M. "Cooperative gaze holding in binocular vision", *IEEE Control Systems*, pp. 24-33, 1991.
- [28]. Cova A.C. and Galiana, H.L. "Bilateral control of vergence and accommodation.", Proceedings of the 17th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1453-1454, 1995.
- [29]. Dallos, P. J., Jones, R. W. "Learning behavior of the eye fixation control system.", *IEEE Trans. Autom. Control.* AC-8:218-27, 1963.
- [30]. Daniilidis, K., Krauss, C., Hansen, M., and Sommer, G. "Real-time tracking of moving objects with an active camera", *Real-Time Imaging*, vol. 4, pp. 3-20, 1998.
- [31]. Desimone R., Wessinger M., Thomas L., and Schneider W. "Attentional control of visual perception: cortical and subcortical mechanisms.", *Cold Spring Harbor on Quantitative Biology, Vol. LV: The Brain, Cold Spring Harbor Laboratory Press*, pp. 963-971, 1990.
- [32]. Dursteler, M., & Wurtz, R., "Pursuit and optokinetic deficits following chemical lesions of cortical areas MT and MST." *Journal of Neurophysiology*, 60(3), 940–965, 1988.
- [33]. Eliasmith, C. and Anderson C.H., "Neural Engineering." MIT Press, 2003.
- [34]. Fiala, J.C, Wavering, A.J. "On visual pursuit systems", NISTIR 5513, National Institute of Standards and Technology, Gaithersburg, MD, 1994.
- [35]. Galiana H.L. and Outerbridge J.S. "A bilateral model for central neural pathways in vestibuloocular reflex.", J. Neurophsyiol., 51 (2): 210-241, 1984.
- [36]. Galiana, H.L. "A new approach to understanding adaptive visual- vestibular interactions in the central nervous system.", *Journal of Neurophysiology* 55: 349-374, 1986.
- [37]. Galiana, H. L. "A nystagmus strategy to linearize the vestibule-ocular reflex.", *IEEE Transactions on Biomedical Engineering*, vol. 38, no. 6, 1991.

- [38]. Gardner J. and Lisberger S. "Linked target selection for saccadic and smooth pursuit eye movements.", J. Neurosci., 21 (6): 2075-2084, 2001.
- [39]. Helmholtz, H. von. "Handbuch der Physiologischen Optik." *Leipzig: Voss.* 874 pp., 1866.
- [40]. Hendry SH. and Reid C., "The koniocellular pathway in primate vision.", Annual Review Neuroscience, Vol. 23, pp. 127-153, 2000.
- [41]. Henn, V., Young, L. R., and Finley, "C. Vestibular nucleus units in alert monkeys are also influenced by moving visual fields.", *Brain Res.* 71: 144–149, 1974.
- [42]. Hoffman, K.P., and Distler, C., "Quantitative analysis of visual receptive fields in neurons in nucleus of the optic tract and dorsal terminal nucleus of the accessory optic tract in monkeys.", *Exp. Brain Res.* 69: 635–644, 1989.
- [43]. Hubel, D. H., & Wiesel, T. N. (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.", *J Phys.-London*, 160: 106-154, 1962.
- [44]. Humphreys G.W. and Miller H.J. "Search via recursive rejection (SERR): a connectionist model of visual search", *Cognitive Psychology*, 25(1): 43-110, 1993.
- [45]. Itti L., Koch C., Niebur E., A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (11), 1254–1259, 1998.
- [46]. Itti L.. "Models of Bottom-Up and Top-Down Visual Attention.", *PhD thesis*, California Institute of Technology, USA, 2000.
- [47]. Itti L. and Koch. Ch. "Computational modeling of visual attention.", Nature Reviews Neuroscience, Vol. 2, No. 3, pp. 194-203, 2001.
- [48]. Kandel, E.R., Schwartz, J.H. and Jessell, T.M., "Principles of Neural Science." 3rd Edition. New York: Elsevier Science Publishing Company. Inc., 1991.
- [49]. Kawano, M., Shidara, Y., Watanabe, Y., & Yamane, S. "Neural activity in cortical area MST of alert monkey during ocular following responses.", *Journal of Neurophysiology*, 71(6), 2305–2324, 1994.
- [50]. Keller, E. L. "Participation of medial pontine reticular formation in eye movement generation in monkey." *J. Neurophysiol.* 37:316-32, 1974.

- [51]. Klier, E.M., Wang, H. and Crawford , J.D. "The superior colliculus encodes gaze commands in retinal coordinates." *Nature Neuroscience*, Vol.4 No.6: 627-632, 2001.
- [52]. Koch Ch. and Ullman S., "Shifts in selective visual attention: Towards the underlying neural circuitry.", *Human Neurobiology*, Vol. 4, pp. 219-227, 1985.
- [53]. Krauzlis, R.J., Basso, M.A., and Wurts, R.H. "Shared motor error for multiple eye movements." *Science* 276: 1693-1695, 1997.
- [54]. Krauzlis RJ and Miles FA. "Role of the oculomotor vermis in generating pursuit and saccades: effects of microstimulation." *J Neurophysiol* 80: 2046–2062, 1998.
- [55]. Kuffler,S.W. "Neurons in the retina : organization, inhibition and excitatory problems." *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 17, p. 281-292, 1952.
- [56]. Kuffler S.W. "Discharge Patterns and functional organization of mammalian retina.", *J.Neurophysiology* 16: 37 -68, 1953.
- [57]. Kuniyoshi, Y., and Rougeaux, S. "A humanoid vision system for interactive robots", Proc. Of the First Asian Symposium on Industrial Automation and Robotics, pages 13-21, Bangkok, Thailand, 1999.
- [58]. LaBerge D., Carter M., and Brown V. "A network simulation of thalamic circuit operations in selective attention.", *Neural Computation*, Vol. 4, pp.318-331, 1992.
- [59]. Langer, T., Fuchs, A. F., Scudder, C. A., and Chubb, M. C., "Afferents to the flocculus of the cerebellum in the rhesus macaque as revealed by retrograde transport of horseradish peroxidase." J. Comp. Neurol. 235: 1–25, 1985.
- [60]. Lee J. W., "Prediction in a Biomimetic Controller for Binocular Target Pursuit on a Free Head.", *Master Thesis*, McGill University, Montreal, 2004
- [61]. Levine M.D., "Vision in man and machine." McGraw-Hill, 1985.
- [62]. Lisberger, S. G., Morris, E. J., and Tychsen, L. "Visualmotion processing and sensory-motor integration for smoothpursuit eye movements." *Annu. Rev. Neurosci.* 10: 97–129, 1987.
- [63]. Louw DF, Fielding T, McBeth PB et al. Surgical Robotics: A Review and Neurosurgical Prototype Development. *Neurosurg* 54:525-537, 2004.
- [64]. Marr D., "Vision." Freeman Publishers, 1982.

- [65]. Melvill Jones, G., Milsum, J. H., Spatial and dynamic aspects of visual fixation. *IEEE Trans. Biomed. Eng.* BME-12:54-62, 1965.
- [66]. Milanese R., "Detecting Salient Regions in an Image: from Biological Evidence to Computer implementation.", *PhD thesis*, Dept. of Computer Science, University of Geneva, Switzerland, 1993.
- [67]. Milanese R., Pun T., and Wechsler H., "A non linear integration process for the selection of visual information. Intelligent Perceptual Systems: New Directions in Computational Perception.", *Lecture Notes in Artificial Intelligence*, Springer Verlag, Vol. 745, pp. 322-336, 1993.
- [68]. Missal M, De Brouwer S, Lefevre P, and Olivier E. "Activity of mesencephalic vertical burst neurons during saccades and smooth pursuit." *J Neurophysiol* 83: 2080–2092, 2000.
- [69]. Mumford D. "On the Computational Architecture of the Neocortex, I: The role of the thalamo-cortical loop.", *Biological Cybernetics*, 65, pp. 135-145, 1991.
- [70]. Niebur E. and Koch Ch. "Control of selective visual attention: Modeling the where pathway.", *Advances in Neural Information Processing Systems*, Vol. 8, pp. 802-808, 1996.
- [71]. Olshausen B., Anderson C.H., and Van Essen D.C. "A neural model of visual attention and invariant pattern recognition." *California Institute of Technology, Computation and Neural System Program, CNS Memo* 18, 1992.
- [72]. Olshausen B.A., et.al. "A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information.", *The J. of Neuroscience*, 13(1):4700-4719, 1993.
- [73]. Palatzoukas, D. "Gaze control fusing eye-head coordination with unified saccadic and smooth pursuit modes of operation", *Master Thesis*, McGill University, Montreal, 1999.
- [74]. Privitera C. and Stark L. "Algorithms for defining visual regions-of-interest: Comparison with eye fixations." *Pattern Analysis and Machine Intelligence* (*PAMI*), Vol. 22, No. 9, pp. 970-981, 2000.
- [75]. Posner M., "Orienting of attention." Q. J. Exp. Psychol. 32 (1), 3-25, 1980.

- [76]. Posner M.I. and Petersen S.E. "The attention system in human brain." Annual Review of Neuroscience, Vol. 13, pp. 25-42, 1990.
- [77]. Rashbass, C., "The relationship between saccadic and smooth tracking eye movements." J. Physiol. London 159: 326-38, 1961.
- [78]. Robinson D. A. "The mechanics of human smooth pursuit eye movement." *J. Physiol. London* 180:569-91, 1965.
- [79]. Robinson D.A., "The oculomotor control system: a review." *Proc.IEEE* 56:1032-49, 1968.
- [80]. Robinson D. A. "Cerebellectomy and the vestibulo-ocular reflex arc." *Brain Res.* 71:215-24, 1974.
- [81]. Robinson, D.A. "Adaptive gain control of the vestibulo-ocular reflex by the cerebellum.", *J. Neurophysiol* 39: 954-969, 1976.
- [82]. Robinson, D.A. "Control of eye movements." The Handbook of Physiology The Nervous System II, 1275 – 1320,1980.
- [83]. Robinson, D.A. "The use of control systems analysis in the neurophysio-logy of eye movements.", Annual Review of Neuroscience, 4, 463-503, 1981.
- [84]. Robinson, D. A., Gordon, J. L., and Gordon, S. E. "A model of the smooth pursuit eye movement system.", *Biol. Cybernet*.55: 43–57, 1986.
- [85]. Roelfsema R., Lamme V., Spekreijse H., Object-based attention in the primary visual cortex of the macaque monkey, *Nature* 395 (6700), 376–381, 1998.
- [86]. Rougeaux, S., and Kuniyoshi, Y. "Velocity and disparity cues for robust real-time binocular tracking", Computer Vision and Pattern Recognition Conference (CVPR), pp. 1-6, 1997.
- [87]. Sagi D., Julesz B., Enhanced detection in the aperture of focal attention, *Nature* 321, 693–695, 1986.
- [88]. Sasaki, K., "Electrophysiological studies on oculomotor neurons on the cat.", Jpn. J. Physiol. 13:287-302, 1963.
- [89]. Schluns, K., Fellenz, W., Koschan, A., and Tescner, M. "A modular 10-DOF vision system for high-resolution active stereo", *Technical Report CITR-TR-2*, ISSN 1171-7483, University of Auckland, New Zealand, March 1997.
- [90]. Scientific American, pp.58, May 2005.

- [91]. Sejean, J. "Implementation of a biologically relevant gaze controller using sensory fusion and shared motor drives" *Master Thesis*, Biomed. Eng. McGill University, Montreal, QC, 2003.
- [92]. Shulman G., Wilson J., Spatial frequency and selective attention to spatial location, *Perception* 16 (1), 103–111, 1987.
- [93]. Skavenski, A. A., Robinson, D. "A.Role of abducens motoneurons in the vestibuloocular reflex.", J. Neurophysiol. 36:724-38, 1973.
- [94]. Stratton, G.M., "Symmetry linear illusion, and the movements of the eye.", *Psychological Review*, 12(2), 82-96, 1906.
- [95]. Suzuki, D. A., and Keller, E. L. "Visual signals in the dorsolateral pontine nucleus of the monkey: Their relationship to smooth pursuit eye movements.", *Exp. Brain Res.* 53: 473–478, 1984.
- [96]. Takagi M, Zee DS, and Tamargo RJ. "Effects of lesions of the oculomotor cerebellar vermis on eye movements in primate: smooth pursuit.", *J. Neurophysiol* 83: 2047–2062, 2000.
- [97]. Thorpe S.J., Fize D., and Marlot C., "Speed of processing in the human visual system.", *Nature*, Vol. 81, No. 6582, pp. 520-522, 1996.
- [98]. Tian JR and Lynch JC. "Corticocortical input to the smooth and saccadic eye movement subregions of the frontal eye field in Cebus monkeys." J. Neurophysiol 76: 2754–2771, 1996.
- [99]. Treisman A. and Gelade G. "A feature integration theory of attention.", *Cognitive Psychology*, 12: 97-136, 1980.
- [100]. Treue, "Neural correlates of attention in private visual cortex.", Trends in Neurosciences, 24, 295 – 300, 2001
- [101]. Tsotsos J.K., "Analyzing vision at the complexity level.", *Behavioral and Brain Science*, Vol. 13, pp. 423-469, 1990.
- [102]. Tsotsos J.K., Culhane S.M., Wai W.Y.K., Lai Y.H., Davis N., and Nuflo F.,
 "Modeling visual attention via selective tuning." *Artificial Intelligence*, Vol. 78, pp. 507-545, 1995.

- [103]. Wagner, R. "Binocular Coordination: Providing stable tracking and rapid reorientation using a bilateral controller inspired by nature" *PhD Thesis*. Biomed. Eng. McGill University, Montreal, QC, 1995.
- [104]. Wavering, A.J., Scheiderman, H, and Fiala, J.C. "High-perfromance tracking with TRICLOPS", Proc. Second Asian Conference on Computer Vision, ACCV, Singapore, 1995.
- [105]. Westheimer, G., "Mechanism of saccadic eye movement." *AMA Arch. Ophthalmol.* 52:710-24, 1954.
- [106]. Wilson, V.J. and Melvill, G., Mammalian Vestibular Physiology, Plenum Press, New York, 1979.
- [107]. Widmayer E.P., Raff H., Strang K.T., Vander's Human Physiology, 10th Edition, McGraw – Hill, pp. 193-95, 2006.
- [108]. Wolfe J.M., "Guided search 2.0: a revised model of visual search." *Psyonomic Bulletin and Review*, 1(2):202-238, 1994.
- [109]. Yasui, S., Young, L. R. "Perceived visual motion as effective stimulus to pursuit eye movement system." *Science* 190:906-8, 1975.
- [110]. Young, L. R., Forster, J. D., Van Houtte, N. "A revised stochastic sampled data model for eye tracking movements." *Proc. Fourth Annual NASA-University Conference on Manual Control*, NASA SP-192:489-509, 1968.
- [111]. Young, L. R. "Pursuit eye tracking movements." Control of Eye Movements, New York, pp 429 – 443, 1971.
- [112]. http://www.ahfmr.ab.ca/publications/newsletter/Winter05/www.files/inside/neuro arm.htm
- [113]. http://www.berkshirecc.edu/business/plasticsCurriculum/PIM06_files/image029.j pg
- [114]. http://www.computermotion.com
- [115]. http://www.intuitivesurgical.com/index.aspx
- [116]. http://psych.athabascau.ca/html/Psych402/Biotutorials/
- [117]. http://sky.bsd.uchicago.edu/lcy_ref/synap/visualpath.html#chiasma
- [118]. http://world.honda.com/ASIMO/