University of Genova - Faculty of Engineering
Department of Communication Computer and Systems Science

LIRA - Lab Laboratory for Integrated Advanced Robotics

Integration of Inertial and Visual Information in Binocular Vision Systems

Francesco Panerai

Doctoral Degree Thesis
Robotics, X° Course

Thesis Supervisor: **Prof. Giulio Sandini**

Course Coordinator: **Prof. Piero Morasso**

Acknowledgments

First of all, let me thank my advisor, Giulio Sandini. Nothing of this work would have been possible, without his guidance, in many senses. He has been always source of inspiration, and from time to time, terrific and tireless co-author. I have have learned a great deal from him...at least I think!

Part of this work was developed in collaboration with Carla Capurro; she has been pioneer in the dynamic control of vergence and deserve a special thank to have provided partial foundations to the framework of the present work. I must also thank Emanuele Castellazzi and Paolo Casanova for their numerous "hardware offspring"; they have been very patient during the soldering of micro resistors and micro capacitors.

Many other people have helped and contributed to make LIRA Lab a very pleasant place during these three years: Enrico Grosso and Massimo Tistarelli, Massimo Massa and Francesco Buemi, Paolo Questa, Giorgio Metta, Fabio Berton, Paolo Rossi, Laura Orsini, Matteo Perrone, Daniela Ferralasco and Roberta Ferrari. I am sure someone is not on the list; nevertheless, he should consider himself included.

I need to thank also my parents: support and love have never been missing.

My final thanks is reserved to Carla: her friendship is irreplaceable, her smile always present.

Abstract

This thesis deals with part of the whole problem of acquiring and maintaining stable vision. More specifically it focuses on the integration of inertial and visual information for gaze stabilization. The approach is biologically inspired, and aims to provide a robot system moving in the environment with visually stable incoming images. Stabilization of gaze is a fundamental requirement of an active visual system for at least two reasons: i) it increases the robustness of dynamic visual measures during observer's motion; ii) it provides a reference with respect to the environment [Ballard and Brown, 1992].

Two important key points emerging from this work are: 1) the integration of two different types of sensory information - inertial and visual - extends the performance (or effectiveness) of gaze stabilization to a wider dynamic range of head movements/disturbances, 2) when combined with coordinated control of vergence, gaze stabilization strategies can be refined, adaptively changing the Vestibular Ocular Reflex (VOR) gain to account for the effects of distance and eccentricity. It is in fact with continuous control of convergence/divergence of cameras that the system is able to perform the following tasks: a) to estimate fixation distance and use it as a context signal to modulate VOR gain, b) to exploit a selective depth filter and limit the visual processing of the whole field of motion to the plane of fixation ; this strategy let the system ignore any other competing motion, not useful for image stabilization, c) to deal with asymmetric requirements of eyes dynamics, especially when fixating eccentric targets at short distances (either for Angular VOR and Linear VOR).

The experimental results are based on the use of one binocular vision system equipped with space-variant sensors and a prototype vestibular system. Several components modules have been designed with off-the shelf miniature sensors (piezoelectric vibrating forks and heat sensitive resistors) and are able to measure, when combined together, three perpendicular angular velocities and three linear acceleration. In particular, within the framework of our robotic application, the following issues are discussed: i) the relations between eye-head geometry, fixation distance and stabilization performance; ii) the computational requirements of the visuo-inertial stabilization approach compared to a visual stabilization approach; iii) the evaluation of performance of the visuo-inertial strategy in a real-time monocular and binocular stabilization task.

The results show that the integrated approach is indeed valuable: it makes use of visual computational resources more efficiently, extends the range of motions or external disturbances the system can effectively deal with, and reduces system complexity.

Index

1 Introduction

The world we perceive through our visual sense is a stable world; this percept is so common experience, that at first glance, the above statement might even seems obvious. Stop thinking for a while to static observers with immobile eyes and imagine observers moving through the environment, moving the eyes at the same time. What is likely to be the image on the retina in such complex dynamic situations? As consequence of body and head movements, the image of the external world that falls onto the retina, moves across it; still, in most cases, we perceive the external world as stable.

How does the visual system manage to achieve this goal ? Perceptual stability of the visual world is primarily a question of physical stability of the retinal image. It is known that at retinal level image tremor is always present [Snodderly and Kurtz, 1985] and, to some extent, it is "even required" to increase sensitivity to some visual stimuli (e.g. contrast for low spatial frequency discrimination [Steinman and Levinson, 1990]). On the other hand, the consequences of large amplitude motions of the retinal image are deterioration of acuity, and consequently, of discrimination and recognition abilities. When our visual field is stable, we experience enhanced sensitivity to individually moving objects [Nakayama, 1983] and to motion parallax, which in turn enhances distance and form perception [Rogers and Graham, 1982]

In humans and primate's visual system, stability of retinal image is achieved with a three-steps process. First step: stabilization of the head with respect to body movements; second step: stabilization of the eyes with respect to head movements; third step, suppression of motion sensitivity during fast eye movements [Burr *et al.*, 1983; Robinson and Peterson, 1985; Shioiri and Cavanagh, 1989] and during self-motion[1].

Within our framework we are mostly interested in the second level of the whole process: eyes are contained in the skull and are influenced directly/indirectly by movements of the head. What is the solution exploited by biological systems in this case?

[1] movement of the image due to self-motion alters the contrast sensitivity function (CSF) [Steinman *et al.*, 1985; Berthoz and Droulez, 1982]

Most frequently, the biological solution is a reflex, neural control of eye movements in the direction opposite to head motion. This neural mechanism combines head motion information (i.e. inertial) provided by the vestibular apparatus in the inner ear, and visual information, extracted from "global" and "local" analysis of retinal image motion. The complementary role played by the combination/integration of these two types of information, it is of paramount importance: it enables to achieve the best performance over the wider possible range of head movements or externally imposed disturbances.

In the robotics domain, on the other hand, systems are most likely to move and interact in the environment using vision as an essential sensory modality. One way to facilitate such a challenging task, is to provide these system with some form of image stabilization mechanism, foreseen to be an essential component, especially for those robots that operate in motion throughout an outdoor environment. Stabilization of gaze is believed to be extremely relevant with respect to the followings: i) to increase the robustness of dynamic visual measures, ii) to provide a reference with respect to the environment [Ballard and Brown, 1992], and iii) to enables the system to preserve functionality of high level vision tasks (e.g recognition), irrespective of external disturbances.

This thesis deals with part of the whole problem of acquiring and maintaining stable vision in binocular vision systems. More specifically it focuses on the integration of inertial and visual information as a methodology for doing this. The approach is biologically inspired, and aims to provide a robot system moving in the environment with "visually stable" incoming images.

1.1 Motivation

The construction of a seeing system is a very fascinating and challenging problem, and many different aspects of it have been extensively investigated, both at computational [Aloimonos *et al.*, 1988; Aloimonos, 1993] and practical level [Crowley and Bajcsy, 1992; Sharkey, 1995; Sharkey *et al.*, 1995], [Weiman, 1995]). Some of these aspects are somehow complicated when we try to make a seeing system move throughout the environment.

Gaze stabilization in binocular vision systems, has received relatively small attention up to now. Most of the approaches proposed are based only on the use of visual information (i.e. pure visual stabilization approach). In order to achieve acceptable real-time performance, these approaches require the computation of image velocity information on specialized and extremely expensive hardware, [Coombs and Brown, 1990; Pahlavan *et al.*, 1992]. Surprisingly, in spite of the fact that most natural systems (at least all those systems equipped with oculo-motor plants) rely not only on visually derived data but also on inertial information to maintain image stability, only few observations have been published in the past on the use of inertial sensors in artificial systems [Vieville and Faugeras, 1990; Brooks, 1996] and only few implementations have appeared using inertial information for line-of-sight stabilization (e.g. [Algrain and Quinn, 1993]). On the contrary, use of inertial information in robot navigation has recently received attentions of different groups [Barshan and Durrant-Whyte, 1995; Lobo and Dias, 1997], particularly because of the availability of small, low cost inertial devices.

When a seeing system moves through the environment displacements on the image plane may occur due to "voluntary" movements of the cameras and/or the head, to "external disturbances" (e.g. during motion/navigation on uneven terrain) or to motion of an object in an otherwise stable environment. In most of these cases, visual information represents just one aspect of the "event" and the use of other sensorial sources, when in appropriate contextual situation, may simplify the general solution of the stabilization problem.

The importance of inertial information in image stabilization is evident if, while reading this page, you turn your head to one side. As you do so, your eyes "compensate" by rotating in the opposite direction so that the image of the text remains stable and sharp on your retina. In principle this behavior could be implemented entirely on the basis of visual cues by tracking the part of the image which is currently fixated. Alternatively one could think of using a copy of the motor commands sent to the head, to generate a compensatory motion of the eyes [2]. Finally one could use an inertial sensing device to measure the rotational velocity of the head and exploit this independent source of information (i.e. not relying on vision and/or internal motor commands) to generate compensatory movements. This last solution applies to both voluntary and passive motion of the head and body.

Because of recent technological advances, small size inertial devices are now available at reasonable cost; it has become therefore feasible to combine multiple sensors to obtain complete motion measurements. With three angular sensors and three linear sensors it is possible to measure rotational and translational motion along three perpendicular axes in vestibular-like manner. Even with multiple sensors, compound system of relatively small overall dimensions can be obtained, enough to be easily integrated in a binocular vision system. We have bought standard off-the-shelf inertial devices and developed our custom electronics and, particularly, control strategies.

The use of inertial information combined with vision was thought from the beginning to be promising, at least for three reasons : i) the latency of the inertial sensors in providing motion information is shorter[3] with respect to vision sensors and, consequently, achievable servo rate based on inertial information could, in principle, be higher; ii) by combining vision with inertial sensor data, the amount of

[2]this solution only applies to internally generated motion and does not generalize to external disturbances such as those generated during locomotion or navigation inside a moving vehicle.

[3]latencies are related to the amount of sensor-generated data, much larger in the case of images

vision-dedicated resources can be in principle reduced, and overall computational resources of the robotics system optimized; iii) furthermore, the inertial data do not depend on visual processing (and vice versa) and their integration does indeed add a completely new and independent data source. This is particularly true in case of external disturbances where even the proprioceptive information measuring head rotation is not relevant. In this case inertial information seems the only source of extra-retinal signals and a powerful tool to increase the overall performance of the system.

Last, but not least, gaze stabilization strategies are intrinsically "binocular", in the sense that a binocular systems can exploit "binocularity" to derive distance cues and to optimize visual processing. In fact, the knowledge of fixation distance and the reduction of spatial extent of processing to image areas corresponding to the plane of fixation, are important issues to achieve the goal of building a vision system that is able to adapt its stabilization performances to the required contextual situations.

1.2 Overview

The content of the thesis is outlined here. Chapter 2 forms the background of the gaze stabilization problem from a kinematics point of view. The formalism provides the mean to show the influence of geometric parameters of the vision system (e.g. baseline, eye to neck distance) and fixation distance on the dynamics of compensatory eye movements. Two cases are treated: pure angular rotations and pure linear translations of the vision system. We also shows that, in both cases, eccentric gaze situations require independent, asymmetric control of eye movements, in particular when fixating objects in the near space.

Chapter 3 guides through a revision of basic concepts of the oculo-motor system and vestibular apparatus: eye movements, vestibular morphology and neural pathway of vestibular signals. Data on vestibulo-ocular dynamic performance are presented, particularly those acknowledged in literature for humans; moreover data on the real effectiveness of compensatory eye movements are also given: retinal movement, in fact, is often reduced but never completely removed on the basis of inertial data alone.

In chapter 4 we revised the bio-engineering literature on vestibulo-ocular reflexes (VOR), optokinetic reflexes (OKR), and visual-vestibular interaction (VVI). We derive, on the basis of such modeling our robotic control scheme exploiting visuo-inertial information; we also describe a computational framework to combine stabilization strategies with binocular control of vergence. Important consequences are : i) the possibility to derive distance cues on the basis of vergence information and, ii) to limit image processing of motion field to the plane of fixation. Both these issues are important contributions for a further refinement of the stabilization strategies, for angular and linear reflexes.

Chapter 5 describes the artificial vestibular system, designed and built during this thesis. It is composed of two parts, in analogy with the biological system: rate gyros, used to measure angular velocity and accelerometers used to measure linear motion. The sensors are placed roughly orthogonal to each other so the axis of rotation/translation can be mechanically broken into three components. The system was designed to be modular, as small as possible (realized using surface mounted technology) and easy to interface within a robotics setup.

Chapter 6 describes the experimental part. Most of the experiments have been performed monocularly and using externally imposed rotational stimuli of the head; the data collection process is detailed and the criteria for comparison between the pure-inertial and the visuo-inertial approaches are given. Finally, visuo-inertial stabilization performance is tested in relation to extraction of dynamic visual cues: an index of expansion/contraction (i.e. divergence), is estimated under stabilized and non-stabilized gaze conditions. Comparison of the obtained results, proves indeed that the advantages are quite relevant, particularly in terms of robustness of the visual measure. Discussion and some future research directions are finally given.

2 Kinematics of stabilization

> "It is sometime useful to approach problems top-down, and from knowledge of the functions that a system must perform, deduce what elements or subsystems must necessarily exist within the system in order for it to be able to perform appropriately" [Gregory *et al.*, 1995].

In this chapter we intend to address the problem of stabilization from a rather analytic point of view. Stabilization control determines the correct direction and the appropriate amount of eye movement to avoid blur and loss of acuity, during motion of the head. Our aim is to implement this function in a robot vision system, therefore we have started by asking the following questions: "what exactly (i.e. mathematically) looks like the expression of the eye velocity required to maintain eye fixation on a certain object of interest for a general rotational or translational movement of the head ? Does it give the same result for the two eyes? Is it related to the general eye-head configuration of the vision system ? Does a kinematic description provide us with some insights of properties that must necessarily exist for the robot system to act appropriately ?"

We have answered these questions using a simple geometric model of a binocular system and deriving the kinematics of eye movements to maintain fixation for both, rotations and translations of the head in an horizontal plane.

2.1 Rotational movements

In a binocular vision system the eye velocity required to perfectly stabilize the image of a stationary object when the head rotates depends on geometrical parameters of the eye-head system and on the distance of the fixation point.

In figure 2.1 the schematic geometry of a binocular system (at least for stabilization around a vertical axis) is shown and the relevant parameters are indicated. In particular, both the inter-ocular distance (or baseline) **b**, and the distance **a** between

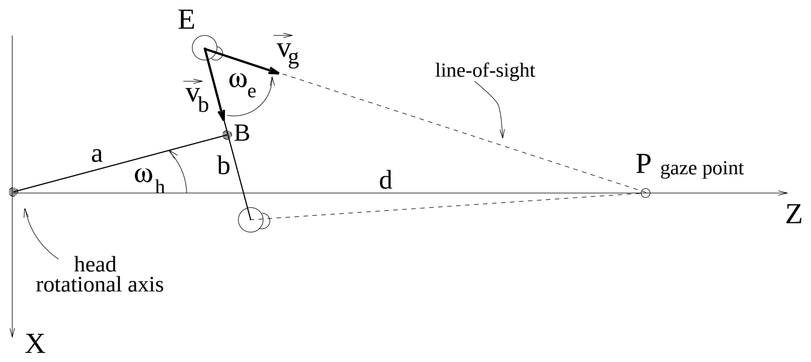


Figure 2.1: Geometry of the eye-head system showing the parameters relevant to inertial and visual measures.

the rotational axis of the head and the baseline, affect the relationship between eye and head velocity. Moreover the eye velocity required to stabilize the image of a stationary object when the head rotates varies inversely with viewing distance **d**.

The analytical relation among these parameters can be derived by considering the kinematics of this model, and imposing the constraint that the eye **E** maintains gaze at point **P** when the head rotates. Consider two vectors, $\vec{v_g}$ and $\vec{v_b}$, laying on the ZX plane and connecting the eye position **E** respectively with gaze point **P** and mid-baseline point **B**.

The instantaneous eye angle ω_e can be expressed according to the following :

$$\omega_e = \arctan\left(\frac{\vec{v_b} \times \vec{v_g}}{\vec{v_b} \cdot \vec{v_g}}\right). \tag{2.1}$$

Substituting the vector symbols with their expression in terms of **a** ,**b** , **d** , ω_h and constraining gaze points to the Z axis, we get:

$$\omega_e = \arctan\left(\frac{d\ \cos\omega_h - a}{\frac{b}{2} - d\ \sin\omega_h}\right). \tag{2.2}$$

Differentiating eq. 2.2 with respect to time gives:

$$\dot{\omega}_e = \left[\frac{d^2 - d\ (\frac{b}{2}\sin\omega_h + a\cos\omega_h)}{\frac{b^2}{4} + a^2 + d^2 - b\ d\sin\omega_h - 2a\ d\cos\omega_h}\right]\ \dot{\omega}_h \tag{2.3}$$

Equation 2.3 determines the relationship between eye velocity $\dot{\omega}_e$ and i) the geometrical parameters of the eye-head system (i.e **b** and **a**) and ii) distance **d** of the fixation point **P**, for any given head velocity $\dot{\omega}_h$. A more general expression for

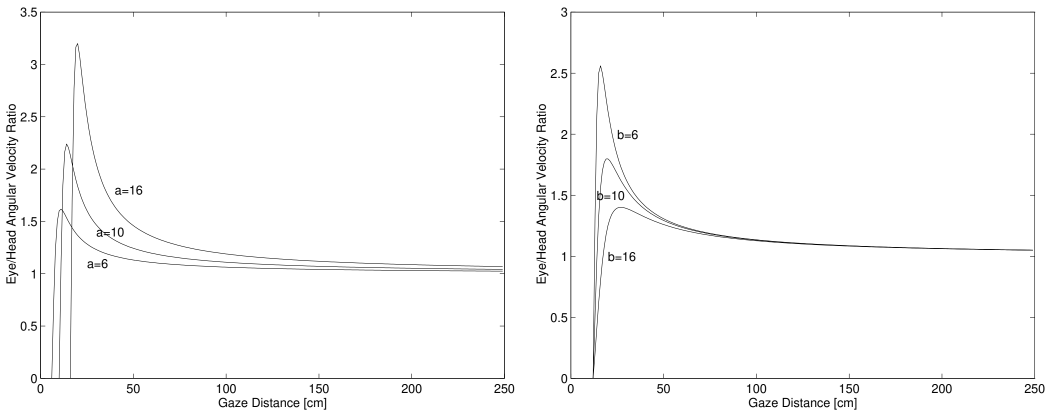


Figure 2.2: Eye-head velocity ratio as a function of distance **d** of the fixation point for different values of **a** and **b** (see geometrical modeling in figure 2.1). Left: baseline **b** is fixed (6 cm) and eye-to-neck distance **a** increases (from 6 to 16 cm). Right: **a** is fixed (12 cm) while baseline **b** increases (6 to 16 cm). Different eye-head configurations requires different dynamics of the eyes with respect to the head.

eccentric gaze points is not considered here but it is easily obtainable (see [Panerai and Sandini, 1997]).

The location of the eyes in the head (defined by **b** and **a**) influences the required eye velocity. In figure 2.2 the ratio between the eye velocity $\dot{\omega}_e$ and the head rotational velocity $\dot{\omega}_h$ required to stabilize the image of an object, is plotted as function of fixation distance **d**. Different choices of the **b** and **a** parameters determines different shapes of eye-head velocity ratio curves. The sample curves give an idea of possible dynamic requirements for existing active vision systems [von Seelen and Madden, 1992], which adopt different geometrical solutions and have different dimensions (i.e. **a** and **b** parameters). With reference to figure 2.2, it is worth noting that the eye velocity $\dot{\omega}_e$ required to maintain fixation on near objects can be as high as twice the value of $\dot{\omega}_h$. In the [25-200] cm range of fixation, the optimally effective amount of oculo-motor compensation needed to obtain gaze stabilization can change rapidly.

Therefore from pure kinematics considerations, it emerges that an important parameter that we might want to consider to adapt rotational reflex oculo motor control to different contextual situations, is distance. In the case of motion of the head, compensatory eye movements generated through a constant gain will not be sufficient to stabilize the visual world; on the other hand, if distance information might be included in the strategy that generate the oculo motor compensation, in principle we could achieve a stabilization of the world very close to ideal. It is not evident, at this stage, how the system can determine distance information, only relaying on its visual capabilities. This issues will be discussed more in details in chapter 4, where

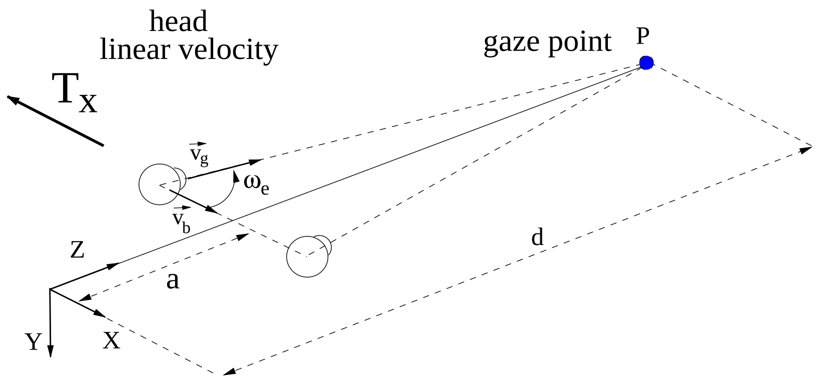


Figure 2.3: Oculo-motor response to a linear interaural translation of the head.

a model on how to compute these information and how to use them for stabilization of rotational movements of the head, will be presented.

2.2 Linear movements

In figure 2.3 a binocular system fixating an object at distance **d** translates at velocity T_x along the X axis (translation along the interaural axis). Using the same formalism used in the previous section, we derive the analytical expression of eye velocity $\dot{\omega}_e$ required to maintain fixation on the object **P** positioned at distance **d**. The vectors $\vec{v_g}$ and $\vec{v_b}$, represent the instantaneous gaze and baseline orientations. The eye angle ω_e can be determined according to the following:

$$\omega_e = \arctan\left(\frac{\vec{v_b} \times \vec{v_g}}{\vec{v_b} \cdot \vec{v_g}}\right). \tag{2.4}$$

If we substitute the vector terms with their expressions as a function of **a**, **b**, **d** and **x** parameters[1], we obtain:

$$\omega_e = \arctan\left(\frac{d-a}{\frac{b}{2}-x}\right). \tag{2.5}$$

Differentiating with respect to time gives:

$$\dot{\omega}_e = \left[\frac{(d-a)}{(\frac{b}{2}-x)^2+(d-a)^2}\right] \cdot T_x \tag{2.6}$$

Eq. 2.6 defines the relationship between eye velocity $\dot{\omega}_e$ and translational velocity T_x and accounts for viewing distance (**d**) and lateral translation (x). Figure 2.4 represents graphically the gain $\dot{\omega}_e \setminus T_x$ with respect to distance **d**. It is worth noting that

[1]lateral translation with respect to gaze point.

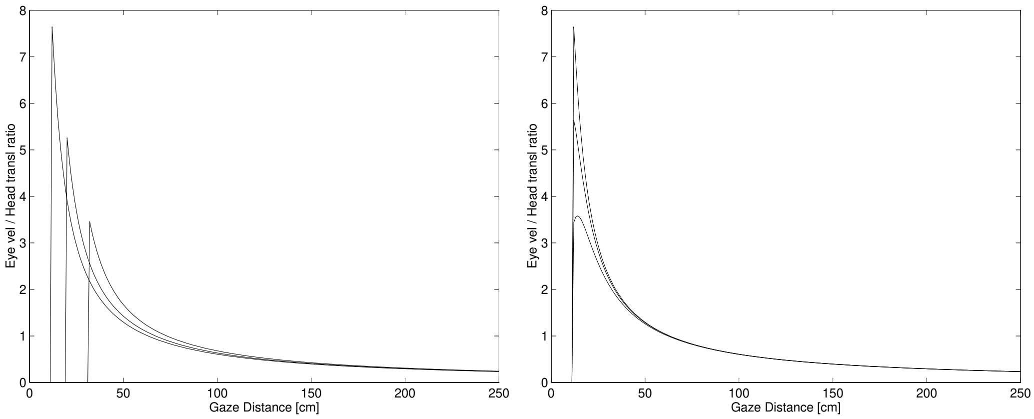


Figure 2.4: Velocity ratio as a function of distance **d** of the fixation point for different values of **a** and **b**). Left: baseline **b** is fixed (6 cm) and eye-to-neck distance **a** increases (from 6 to 16 cm). Right: **a** is fixed (6 cm) while baseline **b** increases (6 to 16 cm).

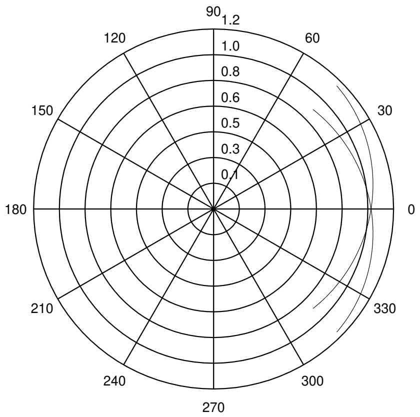
the oculo-motor compensation can be quite demanding for objects at close fixation (i.e. [20-150cm]), but decreases inversely with distance; therefore, an object at infinite distance do not require, in principle, any oculo-motor compensation irrespective of how fast translation is occurring.

Once more the kinematics analysis has provided evidence that distance information is extremely relevant and should be considered if correct oculo-motor compensation is to be achieved. Moreover, when the eye is fixating an object in the range [50-200] cm, the gain (i.e. required eye velocity over linear translation) changes dramatically: from a value of 0.5 at a distance of 150 cm, the gain raises to about 2.0 at 50 cm. In terms of ocular velocity, this means that changing fixation from a point at 150 cm distance to another at 50 cm, while translating at velocity T_x, requires to increase by four times the eye velocity if fixation is to be maintained.

These considerations suggest the necessity of a mechanism tuned to distance information in order to generate compensatory camera movements in response to linear translation.

2.3 Eccentric gaze

When looking an eccentric target in the near space, binocular compensatory eye movements require, at least in theory, asymmetric motor responses [Hine and Thorn, 1987]. From a pure kinematics point of view, the origin of this asymmetric requirement results clear if one compares the analytical expressions of angular velocities of



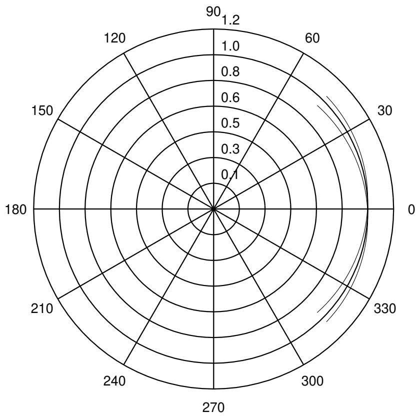


Figure 2.5: Polar plots of the gains of the two eyes during binocular fixation of a target: fixating a target at a distance of 30 cm (left), and at a distance of 70 cm (right). It is evident how the optimal gains vary with head position or equivalently with eccentricity of gaze.

left and right eye during rotation of the head while maintaining fixation. The two required eye angular velocities are:

$$\dot{\omega}_{eleft} = \left[\frac{d^2 - d\ (\frac{b}{2} \sin\omega_h + a\cos\omega_h)}{d^2 + \frac{b^2}{4} + a^2 - b\ d\sin\omega_h - 2a\ d\cos\omega_h} \right]\ \dot{\omega}_h \tag{2.7}$$

for the left eye, and

$$\dot{\omega}_{eright} = \left[\frac{d^2 - d\ (\frac{b}{2} \sin\omega_h - a\cos\omega_h)}{d^2 - \frac{b^2}{4} + a^2 + b\ d\sin\omega_h - 2a\ d\cos\omega_h} \right]\ \dot{\omega}_h \tag{2.8}$$

for the right eye. The two expressions presents in numerator and denominators terms with different signs and it is due to the symmetric eye positions with respect to the center of the baseline. Figure 2.5 represents in polar coordinates the gain functions $\dot{\omega}_e \setminus \dot{\omega}_h$defined in equations 2.7 and 2.8. The gains are plotted with respect to head angular position ω_h and for a given distance of fixation $\mathbf{d}$. In the case of an object at 30 cm distance (left), an angular deviation from the frontal direction of 30 deg introduces a relative gain difference between the two eyes of about 0.2; for example, with a head velocity of 200 deg/s, deviating of 30 deg from frontal direction, gives a relative angular differential velocity of 40 deg/s. Such example show that the angular velocities of the two eyes, can be in the near space, rather different. Although in humans there is no clear evidence that during compensatory eye movements binocular fusion is strictly maintained [Westheimer and McKee, 1975], for a robot vision system, on the other hand, this constraints might be more important, especially if the system controls camera movements on the basis of binocular visual cues [Capurro *et al.*, 97].

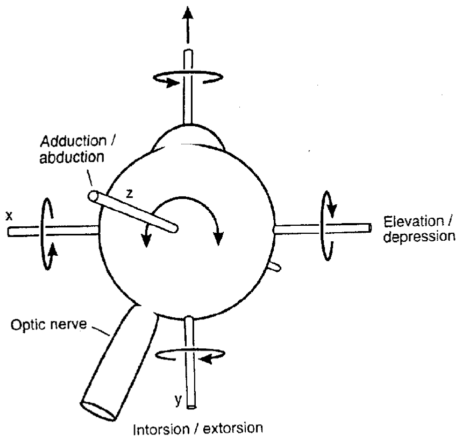
3 Biological Preliminaries

This chapter reviews elements of the oculo-motor system and essential characteristics of the different types of eye movements. The vestibular system is thereafter described together with morphological aspects relevant to define its functionality and performance. The neural pathway, carrying information from peripheral sensory organs (i.e. vestibular) to oculo-motor nuclei, is also considered. It is important to outline that, although vestibular reflexes are known to map directly sensory signals into motor responses, the neural signal goes through important relay stations, where it is either elaborated or modulated, to optimally respond to contextual situation and physical constraint derived from the way the muscle system works. Finally, results derived from psychophysics studies are presented, to asses the performance of the vestibular system during natural activities (e.g. jumping, running, walking).

3.1 The oculo-motor system

The oculo motor system function is to provide the central nervous system with adequate visual sensory information. To achieve this goal, it has developed the ability to either switch or combine five different and complementary types of motor control strategies. In humans, both eyes are moved within their orbital positions by three pairs of muscles (see figure 3.1 - right). Two pairs of muscles are of rectus type, enabling the eye to rotate around the horizontal and vertical axes, the third pair is of oblique type, enabling cyclotorsional movements [Kandel *et al.*, 1991]. To a good approximation, eye movements can be described as rotations around three axes, intersecting in a single point fixed in the orbit. We can distinguish amongst horizontal, vertical and torsional rotations (see figure 3.1 - left).

Important for each foveated visual system, is the ability to bring the central part of the retina quickly onto objects of interest. These types of quick redirections of eye orientation are called **saccades**. It is using saccades that we can catch-up rapidly targets appearing in our peripheral visual field. The velocity of a saccade is determined by the distance of the target in retinotopic coordinates; in certain cases it is not



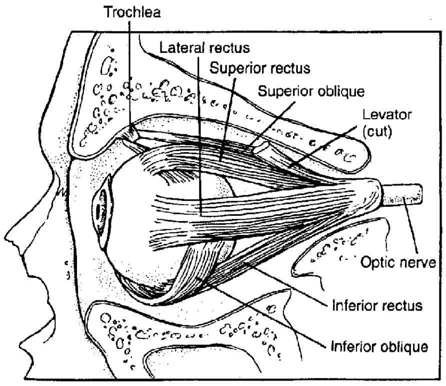


Figure 3.1: Eye movements and eye muscles: movements can be modeled as rotations around three orthogonal axes: azimuth, elevation and cyclotorsion (left), and are obtained by coordinated control of respectively rectus muscle types - azimuth, elevation - and oblique muscle type - cyclotorsion (from Kandel and Schwartz, 1991).

difficult for the oculo-motor system to reach velocities of up to 900 deg/s. Because of their fast execution, no chance for visual correction of trajectory exists once the eye movements started; for this reason, saccades are termed "ballistic eye movements".

Once a target has been reached, it is important to maintain fixation on it, without having to worry of external disturbances (e.g. observer can be in motion) or sudden target motion. To deal with these two situations, two types of eye movements occur: the **vestibulo-ocular reflex** coupled with **optokinetic reflex** - to compensate for head movements - and the **smooth pursuit** - to keep target continuously in the fovea. The basic purpose of vestibulo-ocular and optokinetic mechanisms is to allow us to see clearly even when we are moving. During head movements in any direction, the sensory system contained in the inner ear, the vestibular apparatus, measures accurately the dynamics of transitory head motions. To a first approximation, this information is passed to the oculo-motor system that generate compensatory eye movements in equal and opposite direction, with the overall effects of canceling the retinal slip that would otherwise disturb clear perception of external world. The optokinetic system, on the other hand, contributes to stabilization of the external world, when full-field motion is detected. In such a situation, it generates an appropriate signal to drive the eye to reduce retinal slip. The two mechanisms are complementary: the first, relaying on the dynamic characteristics of vestibular information, has short latency and slow decay, the second, based on visual information, has a long latency and a slow buildup.

To deal with the requirement of following an object in motion the oculo-motor

Eye movement	Eliciting stimuli	Characteristics
saccade	attentional cues	voluntary, conjugate, extremely fast (up to 900 deg/s)
vestibular reflex	vestibular signals	non-voluntary, conjugate, short latency and slow decay
optokinetic reflex	whole retinal slip	non-voluntary, conjugate, long latency and slow buildup
smooth pursuit	retinal image motion	voluntary, conjugate, continuos target motion required
vergence	disparity, accommodation, expansion/contraction	non-voluntary, slow

Table 3.1: Types of eye movements: eliciting stimuli and characteristics.

system developed smooth pursuit abilities. Smooth pursuit is a voluntary movement and requires a moving stimulus; it can deal with object movements up to about 100 deg/s.

The last category of eye movements to be mentioned, is **vergence**. Vergence enables us to perceive single objects (i.e. avoiding diplopia) by controlling with *disconjugate* movements each eye in an horizontal plane (horizontal vergence). Convergence or divergence of the eyes give the possibility to fixate objects at different depths.

Table 3.1 gives a summary of eye movements, their eliciting stimuli and their essential characteristics.

3.2 The vestibular apparatus

The vestibular apparatus in humans and animals is operating in every day life to cope with situations often occurring: active movements and brief passive movements of head, trunk and body. It is known that the system does not register sustained movements at constant rate, but it can be argued that this condition rarely occurs in normal life. Its general function is therefore to respond to changes in the rate of motion, although a further distinction envisages three more specific roles, sketched in figure 3.2; these are:

- control of eye movements to stabilize vision during head movements - *vestibulo-ocular reflexes*; responses are evoked by sensory information generated by semicircular canals and otoliths organs, and are used mainly to control extra-ocular

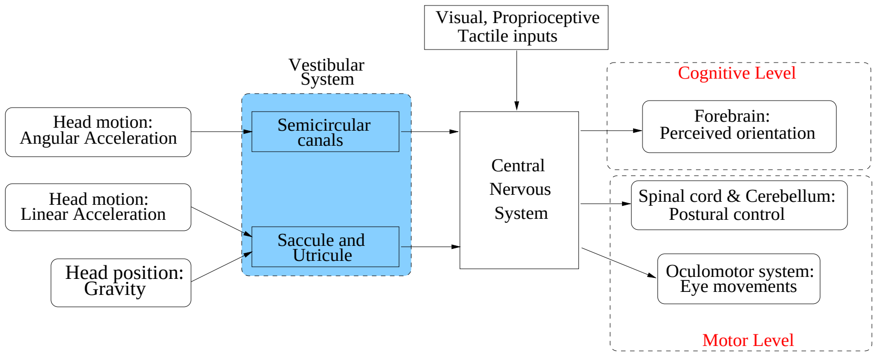


Figure 3.2: Functions of the vestibular apparatus: perceived orientation, postural control, eye movements.

muscles - *vestibulo-ocular reflexes* - and coordinate head and eye movements (i.e. during saccades); the reflex action contributes largely to reduce the movements on the retina, of the image of the object being fixated;

- vestibular information are used in motor control to elicit adjustment of muscle activity and body positions to allow for upright position. These responses are primarily evoked by sensory input from the otolith organs (i.e. utricule and the saccule) which inform the brain about the direction of gravity and linear acceleration in the horizontal and saggital plane. Two of these reflexes act directly on the neck - *vestibulocollic* - and the limbs - *vestibulospinal* reflex;

- the system is the primary end organ of equilibrium, and plays a major role in the sensation of motion and spatial orientation.

3.2.1 Apparatus morphology

The vestibular apparatus in humans is differentiated in two functionally distinguished parts: the three semicircular canals and the two sac-like structures of utricle and saccule, also called otolith organs (figure 3.3).

Each semicircular canal is fluid filled[1] and has a dilatation, called the ampullae, occluded by a deformable structure, the crista-cupula system (see figure 3.4 - right). The general principle sustaining the functioning of the canals is that at each movement of the head, because of the inertia, the fluid contained within canals attempts to retain its original position, exerting a deformating action on the cupula system. Therefore, during angular acceleration, the fluid displaces the cupula system (figure 3.4 - left),

[1]the fluid is called endolymph

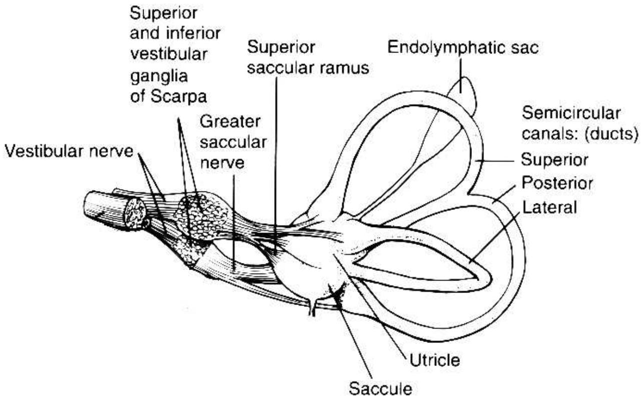


Figure 3.3: Structure of the vestibular apparatus: canals and otoliths organs.

and mechanical deformation causes a change in the discharge rate of the hair cells (i.e. cilia), that are embedded in the cupula itself. The discharge rate turns out to be a monotonic function of this mechanical deformation.

Sensory cells of each canals exhibit a similar directional sensitivity, (see figure 3.5), which is in fact equivalent to a functional polarization: a displacement in one sense produces an increase of firing of the vestibular nerve fibers, a displacement in the opposite direction decreases the firing rate. Although the effective stimulus is the angular acceleration, the hydrodynamics of the canal-cupula-endolymph system[2] are such that a deflection of the cupula tend to be proportional to angular velocity of transient head movements, and not to angular acceleration. Thus canals, are actually velocity transducers [Melvill-Jones, 1971].

The horizontal canals are placed symmetrically in the inner ears (see figure 3.6 - left), and are approximately arranged in three planes orthogonal to one another, forming a three coordinate system (figure 3.6 - right). Such an arrangement enable to detect angular acceleration of the head in any direction in three dimensional space.

As regarding the abilities to measure linear motion, there are two similar structures, contained within the saccule and utricule, called the maculae. Both the saccular macula and the utricular macula are covered by a gelatinous mass called the otolithic membrane, embedding concretions of calcium carbonate, the otoliths. The layout of the membranes and their shape is shown in figure 3.7. These inertial masses make the vestibular organs sensitive to linear acceleration and changes of position of the head in

[2]Note that in man the internal diameter of the canal is small and the fluid flow becomes very heavily damped by viscosity; as a consequence the hydrodynamics of the system performs an integration upon the angular acceleration to which the whole system is exposed.

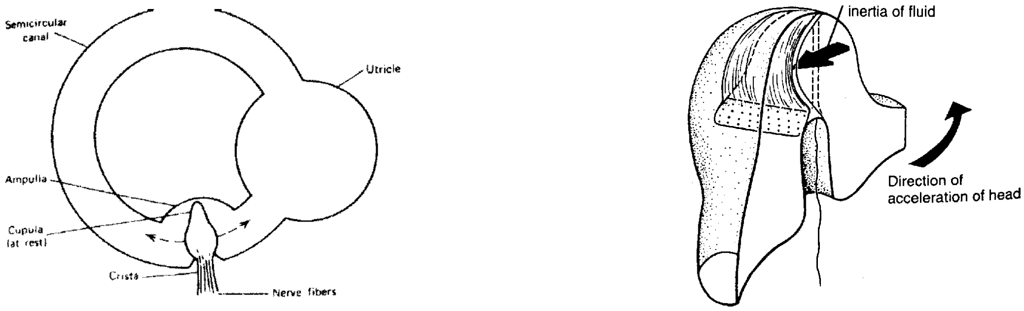


Figure 3.4: Semi circular canal: (right) ampulla, cupula and crista; (left) crista-cupula system is systematically displaced during a rotational acceleration of the head.

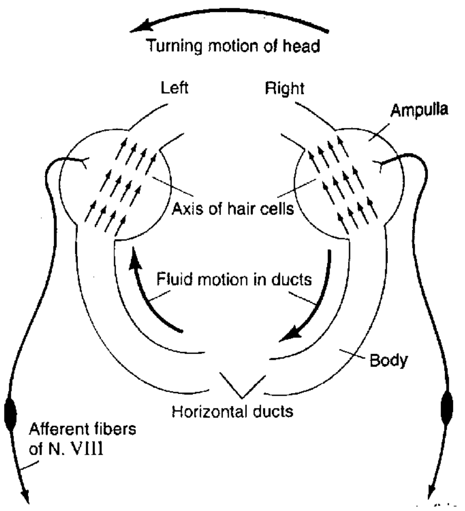


Figure 3.5: Polarization of sensory cell in the canals.

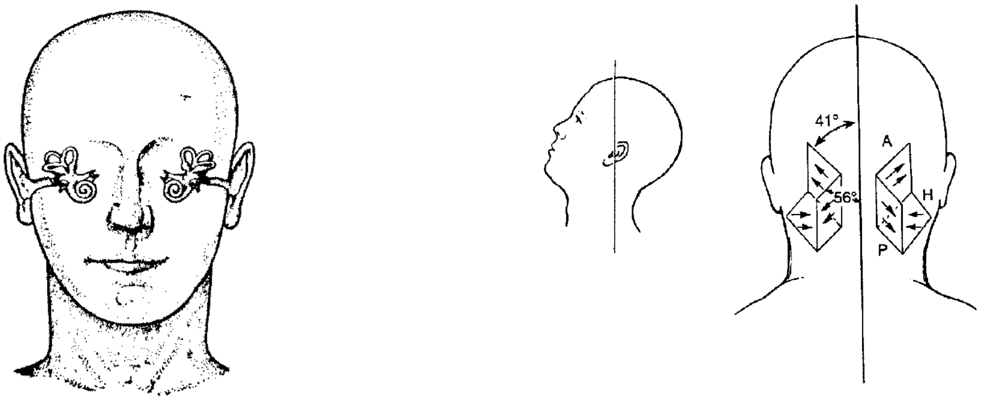


Figure 3.6: Symmetric position of the two vestibular system in the inner ears (left), and arrangement of the three canals onto three orthogonal planes (right).

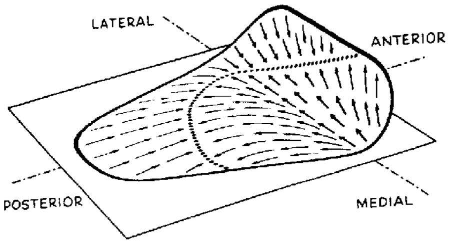
the gravitational field. Hair cells contained in the maculae are oriented in a "fanning" fashion, so that for any direction of otolithic motion, a pair of orthogonal hair cells can be found to supply information about stimulus direction (see directions of polarization in the maculae of the utricule, figure 3.7). Unlike the semicircular canals, the otolith system is not damped by viscous forces, so that the afferent signals reflects, as a first approximation, the linear acceleration experienced by the head [Wilson and Melvill-Jones, 1979].

3.3 Neural pathway of vestibular signals

The organization of the neural pathway from peripheral organs to the ensemble of eye muscles is rather complicated and different for reflexes generated in the horizontal, vertical and cyclotorsional directions. Also, motor nuclei involved in the three different reflexes are rather widely separated. The intent here is to give a simple description of the pathway of the horizontal reflex, but emphasize the functional significance of the different stages, much more important for our purpose.

Vestibular signal originated in the peripheral labyrinth, are first relayed to the vestibular nuclei (see figure 3.8). The main functions of neurons in this nuclei is to setup and distribute excitatory and inhibitory signals to all the oculo-motor nuclei. A second important role of this structure, localized in the medial vestibular nucleus, is to accomplish the integration of canal signal to generate a neural correlate embedding velocity and position information[3] The vestibular nucleus is not the only structure

[3]this important signal is necessary to maintain eye position when the transitory movement infor-



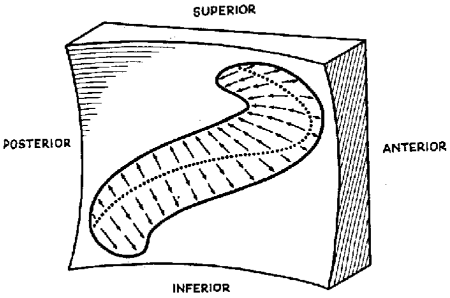


Figure 3.7: Otolith organs: (left) polarization of the saccular macula in the horizontal plane; (right) polarization of the utricular macula in the vertical plane.

Canal stimulated	Muscles contracting	Muscles relaxing
Horizontal	Ips medial rectus (mr)	Ips lateral rectus (lr)
	Contr lateral rectus	Contr medial rectus
Anterior	Ips superior rectus (sr)	Ips inferior rectus (ir)
	Contr inferior oblique	Contr superior oblique
Posterior	Ips superior oblique (so)	Ips inferior oblique (io)
	Contr inferior rectus	Contr superior rectus

Table 3.2: Canal stimulation and muscles contraction coupling.

responsible for the neural integration process; also the cerebellar floculus, and a brain stem nucleus, the nucleus prepositus hypoglossy, are recognized to participates to this function [Keller, 1974].

On the other hand, part of the neural activity occurring in the vestibular nucleus, is relayed directly by short-latency pathways, through monosynaptic connections, to motoneurons in the extraocular motor nuclei [Wilson and Melvill-Jones, 1979]. Of particular interest is also the relationship between canals and eye muscles. It has been shown that artificial stimulation of one canal (e.g. through caloric stimulation) leads mainly to contraction of one muscle in each eye, the so called prime mover [B. Cohen and Bender, 1964], and relaxation of the antagonist. A summary of the principal effects of canal stimulation on eye muscles is given in table 3.2. From a functional point of view, the canals-muscles organization is such that the plane of induced eye rotation is always parallel to the plane of the stimulated canal (see figure 3.9) [Wilson and Melvill-Jones, 1979].

mation generated by canals disappears.

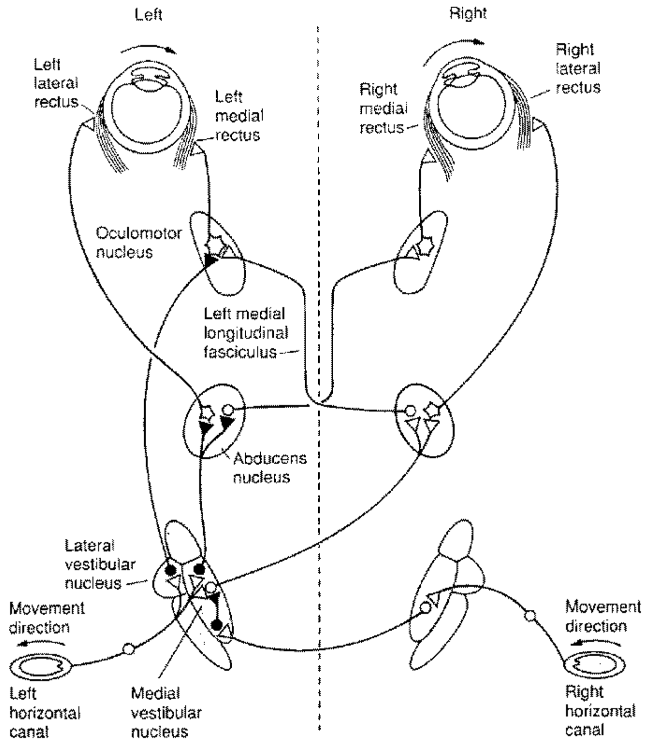


Figure 3.8: Rotations of the head to the left stimulate left canal and inhibits the right canal. The result of the differential discharge is combined at the level of the vestibular nucleus. Excitatory (open symbols) and inhibitory (closed symbols) connections drive in a push-pull fashion the activity of eye muscles. Nucleus prepositus hypoglossi occupies a medial position between medial vestibular nucleus and abducens nucleus.

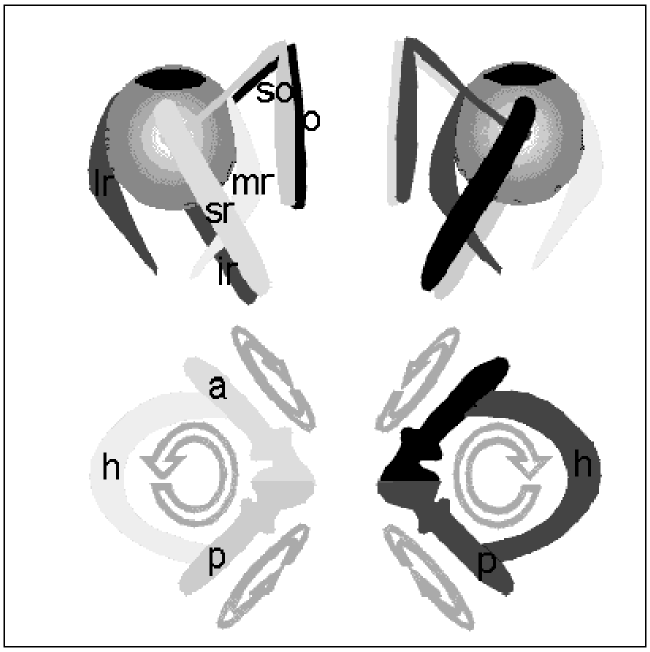


Figure 3.9: Canals and muscle correspondence (from Tweed and Vilis, 1996).

3.4 Range of dynamic measurements

A wide range of active head movements characterizes every day life of animals and humans. Most of them are too short and too fast for *visual-oculomotor* responses to be generated. A suitable example is illustrated in figure 3.10 where highly dynamic eye-head coordination is performed during a saccade to an eccentric target [Morasso *et al.*, 1973]. In 100 ms time, the head has covered almost 18 deg, indicating an approximate velocity of about 180 deg/s, perfectly compensated by the oculo-motor system (see in figure 3.10 direction of gaze (G) after an initial transitory interval, is almost constant).

On the other hand, passive movements are frequently encountered during activities such as walking, running or driving over rough terrain. An analysis of the frequency spectrum of head angular acceleration in these cases reveals that frequency contents are well above the value of 1.0 Hz (see figure 3.11 (bottom)). It is known that visual following above this threshold abruptly deteriorates [Micheal and Melvill-Jones, 1966; Baarsma and Collewijn, 1974]. Figure 3.12 shows gain and phase of eye following movements in man with fixed head [Fuchs, 1967]. Therefore, movements of the head during running or jumping occurs in a frequency range that cannot be compensated using visually-driven oculo-motor responses. On the contrary, figure 3.13 shows that the vestibular system can effectively cope with such high frequency components of

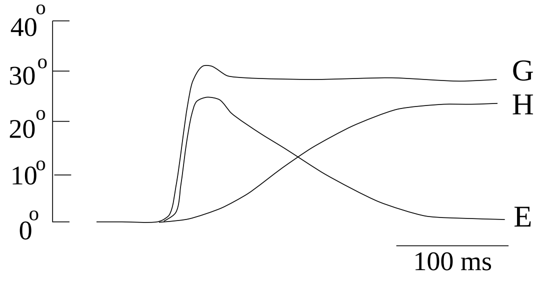


Figure 3.10: Vestibulo-ocular compensation during target acquisition with coordinated eye-head movements: (E) eye position, (H) head position, (G) gaze position.

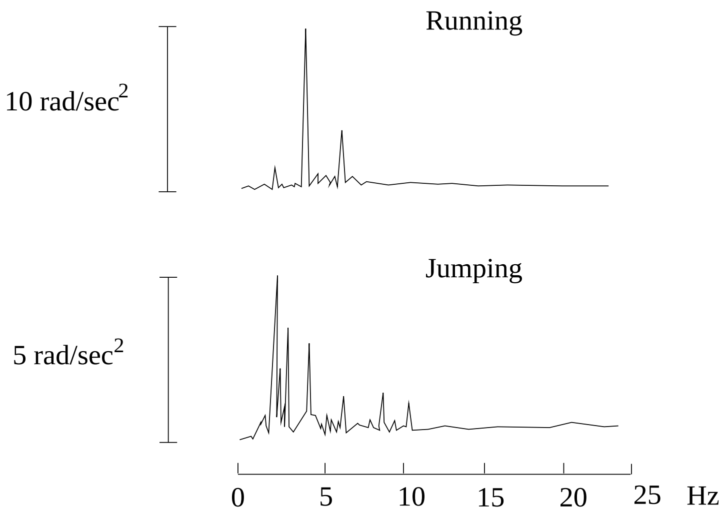


Figure 3.11: Frequency spectrum of human angular acceleration in the saggital plane during running and jumping (from Wilson and Melvill Jones (1979)).

head movements [Keller, 1978]. This experimental data show that the ocular response to rotational head movements remains at a gain of 1.0, at least up to 4-5 Hz.

Other interesting studies have put in evidence that dimensional changes[4] in the canals of different species have the effect to shift the transducing range of angular velocities. In most cases the new range of angular velocity is the most appropriate for the natural frequencies of head movements of an animal of a given size [Melivil-Jones and Spells, 1963; Melvill-Jones, 1969].

3.5 Stabilization performance in natural conditions

Several experimental studies have shown that oculo-motor compensation is mostly less than perfect depending on the size and frequency of body motions. According to some author's opinion, the goal of the oculo motor compensation, is not retinal image stabilization, but rather controlled retinal image motion, adjusted so as to be optimal for visual processing over the full range of natural motions of the body.

This is essentially the position of Skavenski et al. [A.A. Skavenski and Winterson, 1979]. Their experiments show that the gain of the compensatory eye machinery is adjusted to maintain image motion within same range optimal for visual processing. In their opinion, gain is never sufficiently high to produce functional stabilization or sufficiently low to permit images to move too rapidly. The authors agree on the fact that visually guided compensatory eye movements cannot keep up with rapid head or body motion. When such situations occur, the retinal image moves across and acuity suffers. On the contrary, when the vestibular system comes into play acuity is restored. This fact does not mean, in authors' view, that the compensation is perfect. It is in fact recognized [Westheimer and McKee, 1975; Murphy, 1978] that acuity does not suffer until retinal image motion exceeds about $2^o arc/sec$. The experiments conducted by the authors prove that the VOR in combination with visually guided eye movements, fails to completely compensate small amplitude natural head rotations composed of combination of frequencies from d.c. to 7 Hz.

Another study to asses performance of visual-vestibular interaction during active and passive head movements in naturally occurring situations was conducted by Deemer and Viirre [Demer and Viirre, 1996]. They compared ocular responses during actively generated head rotations at 0.8 Hz (i.e. laboratory situation) with response produced during more natural activities: standing, walking and running. The tests were performed under two different optical conditions: wearing 1X spectacles (i.e. normal vision) and 2X, 4X spectacles (i.e. telescopic spectacles). They found that during actively generated head rotations, under normal vision, VOR - in darkness -

[4]Jones and Spell data (1963) show that a horse weighting 450 kg has an internal radius of the endolymphatic canal of 0.19 mm and a radius of curvature of the whole canal of 4.1 mm; for a dog, weighting only 14 kg, the dimensions become respectively, 0.14 mm and 2.4 mm.

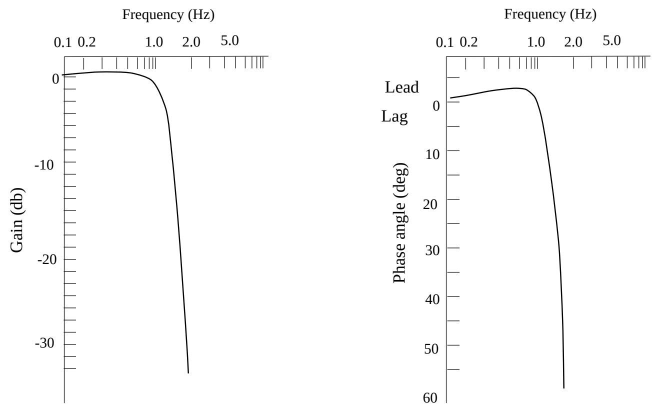


Figure 3.12: Gain and phase of eye following movements when tracking a sinusoidally moving target with fixed head.

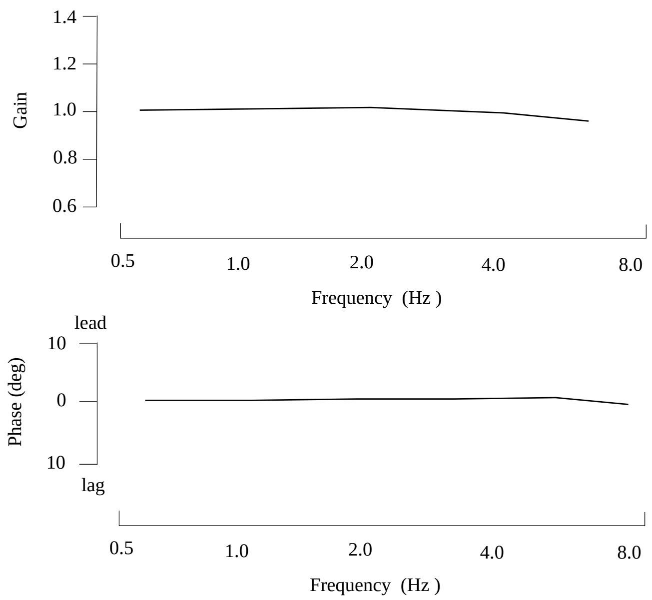


Figure 3.13: Frequency response of ocular compensation during sinusoidal head rotation in the light.

and visually-enhanced VOR (VVOR) have enhanced precision as compared to passively generated head rotation [Demer and Baloh, 1993]. This may suggest that motor efference copy may be employed to augment VVOR during self-generated movements. They also noted that an increase in VVOR gains occurs when wearing telescopic spectacles; however, gains never reach the ideal values needed to prevent retinal slip. This finding was especially true during natural activities, whose high frequency contents were found significant up to 8.2 Hz [G.E. Grossman *et al.*, 1988]. They concluded that although the VOR can generally cope with this range of frequencies [P.D. Poulaski, 1981], as also Hyden [D. Hyden, 1982] and Tomlinson [R.D. Tomlinson, 1980] proved, retinal slip is not completely eliminated, particularly for poorly predictable image motion [G.R. Barnes and Eason, 1987].

4 Modeling Gaze Stabilization

Stabilization of gaze is a solved problem in the majority of biological systems. They can move in the environment using vision among their primary sensory capabilities, without compromising their ability to process visual information. For many animals with mobile eyes, a reflex low level oculo-motor control minimizes the visual effects of body movements or externally induced disturbances. Without such a reflex control, either current fixation, or more in general visual processing, would be definitely undermined. In other interesting cases, such as in the barn-owl or certain types of insects, gaze stabilization is obtained by compensatory head movements [K. E. Money, 1968]. These animals have, in fact, very poor dynamic control at eye level, and supply to this lackness using their flexible neck, to counter rotate the head in response to changes in body orientation.

In vertebrates, it is by combining vestibular and visual information that stabilization of gaze is obtained. The vestibular information is used by the central nervous system in open-loop, that is, the output - an eye movement - does not influence the input - a head movement; visual sensory information, on the contrary, functions in closed-loop; the output - an eye movement - decreases the input signal - the retinal slip velocity. The nature of the these two sensory information - vestibular and visual - is indeed different and the neuronal control mechanism that use them, works optimally for different frequencies and speed of head movements; the first at higher values, the second at lower frequencies and speed values; for this main reason, they are essentially complementary.

Gaze stabilization is a functionality that is foreseen to be important and necessary in many robotics applications. In fact, even more than in biological systems, computer vision systems are very sensitive to image blur effects. Uncontrolled image motion, in many cases, affects the outputs of visual processing algorithms. Therefore, stabilization of line of sight, is particularly useful when the robot system carry onboard one or more visual sensors and moves throughout the environment; undesired motion effects can derive from external causes (e.g. navigation on uneven terrain) or from internal causes (e.g. natural balanced gait movement of the recent walking bipeds).

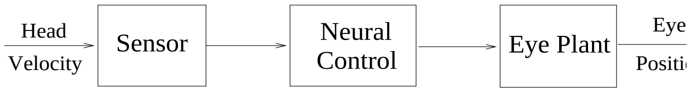


Figure 4.1: Block diagram of the three stages of the vestibulo-ocular reflex (VOR).

The problem of gaze stabilization and the underlying modeling is then interesting for at least two reasons: to understand the way biological systems pertain to such an important visual function, and eventually to give to people interested in building robots, several guidelines on how to realize system that efficiently integrate multi-sensorial information. In particular, the successful realization of a robot system able to stabilize its gaze, will strongly increase the possibility to exploit the visual sensory capabilities to interact with the environment, while moving throughout it.

In the next three subsections, the milestones in the modeling of vestibular and optokinetic reflexes and their interaction will be reviewed. In particular experimental evidence of aspects related to the variability of the gain of the reflex in certain contextual situation (e.g. close distance fixation, synergistic stimulation of angular and linear reflex components) will be treated. The last three subsections describe how part of the modeling aspects reviewed so far have been transferred into a control scheme to stabilize camera movements of a binocular vision system.

4.1 Modeling VOR

To a first extent, the vestibulo-ocular reflex can be conceived as a single open-loop system, that transforms perceived head movements into compensatory eye rotations. It is perhaps the simplest sensori-motor system and a general scheme that reflects the sensori-motor coupling it implements is shown in figure 4.1. A description at different levels of completeness can be given for each of these building blocks: the sensory block - either the otoliths or the semicircular canals - is usually described as a second order system [Mayne, 1965; Young and Meiry, 1966]; the motor block - the eye plant - on the other hand, can be described with a sufficient accuracy by a first order dynamic system. Moreover, the main interest of many researchers studying oculo-motor control, and particularly, gaze stabilization, has been on the central block, processing sensory information to generate appropriate motoneural signals.

In the literature the peripheral vestibular system has been described as a high pass filter - a system that filters out continuos stimuli - with a time constant T_v of about $15-20s$ in man [Young and Oman, 1968]. This value was computed by recording directly oculo-motor responses and it was noted immediately that estimates were significantly higher than the values of semicircular canals time constants measured at

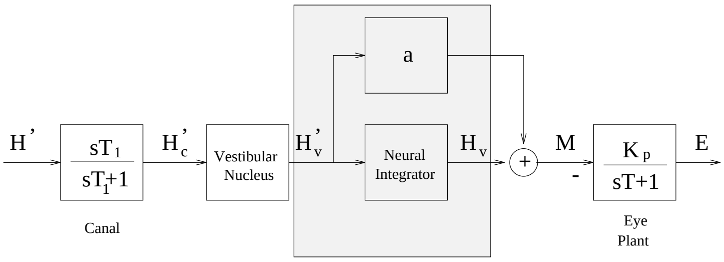


Figure 4.2: Cascaded model of VOR proposed by Robinson. **H'**, **H'c** are respectively head velocity deg/s and canal estimate of head velocity (sp/s); **H'v** is vestibular nucleus estimate of head velocity sp/s, **Hv** is central estimate of head position; M is motoneural firing rate sp/s, **E** is peripheral eye position deg/s.

the level of primary vestibular neurons [Melvill-Jones and Milsum, 1965]. Robinson [Bach-y Rita P., 1971] suggested that some central circuit should exist, which manipulate the canal signals to improve the quality of the ***. Robinson's model (1971) is shown in figure 4.2 and assumes that a central stage provides an estimate of the head velocity. In his view, the central stage contains a neural integrator and it is used to produce a position estimate, or motor efference copy, from the velocity estimate. The central stage output is thus postulated to be the weighted sum of these two estimates.

Further studies concentrated on modeling the effect of adaptation of VOR. This was done by introducing an "adaptation operator" [Young and Oman, 1969],[Malcolm and Melvill-Jones, 1970], that accounts for the response decay to a constant angular acceleration input, in both the peripheral vestibular afferents and vestibulo-ocular reflex. Figures 4.3 and 4.4 show block diagrams of the two models of the VOR containing an adaptation operator; they both have identical VOR transfer function:

$$\frac{\dot{E}}{\dot{H}} = \frac{-sT_c}{sT_c+1} \cdot \frac{sT_a}{sT_a+1} \tag{4.1}$$

where $\dot{E}$ is eye velocity, $\dot{H}$ is head velocity, T_c is the long peripheral vestibular time constant and T_a is the adaptation time constant, but are based on different principles: a negative feedback in Young and Oman model, and negative feed-forward in Malcom and Jones's model. Both of these models describe a VOR response decline following prolonged constant angular acceleration. Nonetheless, they do not address the question of whether the adaptation process occurs in the peripheral or central vestibular structures (i.e. brain stem and cerebellum).

Questions on weather the vestibulo-ocular reflexes could deal with both rotational and translational movements of the head where addressed later by different researchers [U. Schwarz and Miles, 1989; Paige, 1991]. It was found that separate selective mechanisms exist to compensate rotational and translational disturbances of the head.

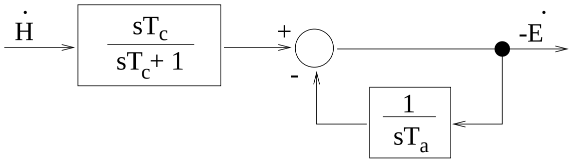


Figure 4.3: Young and Oman (1969) model of VOR adapted response: a negative feedback is here used to create the adaptation operator.

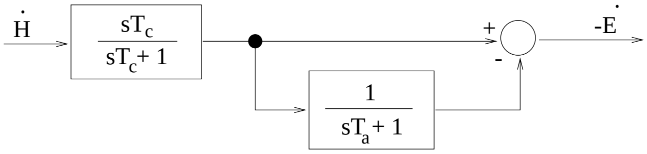


Figure 4.4: Malcolm and Jones model (1970) of VOR adapted response: a negative feed-forward is used to create the adaptation operator.

The general class of VOR responses was thereafter further specified in responses to angular motion of the head, the so called rotational VOR (RVOR), and responses to linear motions of the head, the linear or translational VOR (LVOR or TVOR).

The first model of TVOR, was derived from classical models of the canal-ocular reflex [Bach-y Rita P., 1971], by Buizza, et al.[A. Buizza and Droulez, 1981]. The block diagram is shown in figure 4.5. The authors suggestion was to change only the dynamics of the input mechanical receptors, assuming for the otoliths a description in terms of a first order transfer function. The time constant of the otolith periphery $T_0 = 0.25$ sec, was derived from experimental data while the time constant T_e and the gain K_l were assumed from other models. The model time constants, derived experimentally from recording of input output responses, are nonetheless in contradiction with more direct estimates, obtained from recording at the level of primary

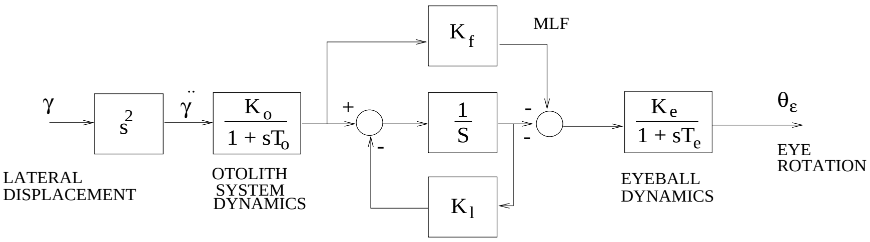


Figure 4.5: Buizza et al. model of the otolith-ocular reflex (OOR).

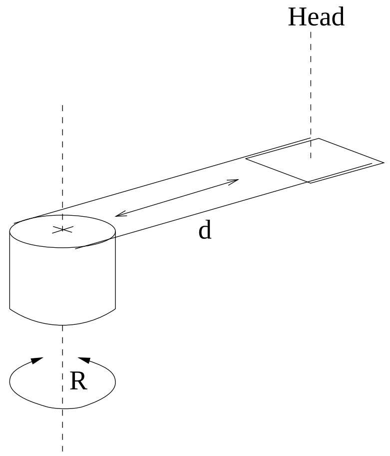


Figure 4.6: Schematic of the rotary apparatus used to stimulate systematically, the otolith and the canals.

vestibular neurons [Fernandez and Goldberg, 1976]. The authors explained that this fact is due to inaccuracies in the representation of the central processes mediating the ocular response, that was derived from others models.

4.1.1 AVOR and LVOR interaction

Ocular responses to rotational and translational movements were always studied separately. Only recently, it was given evidence that **these two reflexes, if properly evocated, combine linearly** [Sargent and Paige, 1991]. Eye movements were recorded in response to horizontal dynamic rotation with head positioned either centrally or eccentrically from axis of rotation (see figure 4.6). When the head was positioned centrally, only rotary vestibular stimuli are applied, these activating AVOR; on the other hand, when the head is placed eccentrically, giving way to an "eccentric rotation", tangential and centripetal linear vestibular stimuli are added to the rotary stimulus, resulting in a combined activation of AVOR and LVOR. Therefore, with the help of such an apparatus, data characterizing the horizontal AVOR alone - canal driven -, the inter aural LVOR alone - otolith driven - and the ocular response to combined AVOR-LVOR stimulation were collected. In addition the experimental apparatus permitted to test two situations, the "nose out" and "nose in" configuration. In the first, LVOR adds synergistically to AVOR, while for the second, the oculomotor responses combine antagonistically, therefore reducing the amplitude of eye compensatory movements. One of the appealing hypotheses emerging from Sargent and Paige's work, is that **the ocular response to complex motions is in fact the summed activity of all component pathways**.

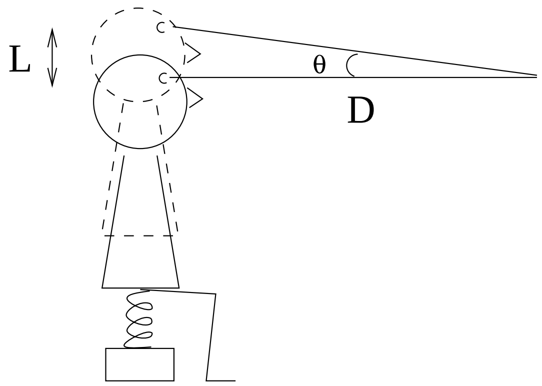


Figure 4.7: Schematic diagram of the experimental setup illustrating linear vertical movements of the subject, and geometric consideration relating to eye movements required to maintain fixation on a target at a given distance D.

4.1.2 LVOR and AVOR dependence upon distance

In a successive study [Paige, 1991] the author showed that the precise amplitude of **VOR response is governed by target distance**. It is then hypothesized that some central nervous system mechanisms exist to determine target distance and dynamically tune the VOR. Paige's experiments consisted in measuring eye response to linear vertical head movements of several human subjects, fixating spot light targets at distances ranging from 33 to 400 cm (see figure 4.7). In geometrical terms, Paige showed that to maintain a stable retinal image of a target at a given distance (D), the ideal vertical angular eye movement (θ) required to compensate for a vertical linear head displacement (L) is given by:

$$\theta = \arctan(\frac{L}{D}) \tag{4.2}$$

It is to say that as target distance increases, the compensatory eye movements are not required to maintain a stable retinal image during linear motion, but it becomes increasingly large as the target moves near.

In repetitive experimental sessions, Paige introduced lenses and prism to induce selective changes in accommodation and vergence. His conclusion was that only the state of vergence strongly influences linear VOR responses. Thus, the state of vergence may provide the brain with a measure of target distance which is then used to bias the angular (canal-mediated) and linear (otolith-mediated) VORs under a variety of motion profiles.

The hypothesis that vergence might provide the necessary internal signal to govern the VOR gain, was also remarked by other investigators; amongst the firsts, Blake-

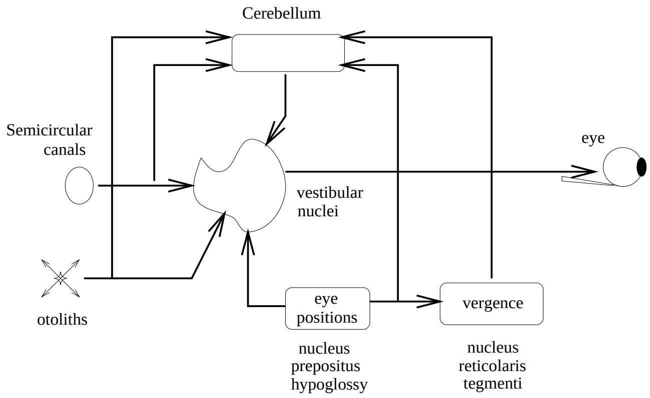


Figure 4.8: Coenen and Sejnoski model: schematic diagram of the processing nuclei and connections considered. Two structures, the nucleus prepositus hypoglossy and the nucleus reticularis tegmenti pontis, are meant to be representative of a class of nuclei in the brain stem carrying eye position or vergence signal.

more and Donaghy, Biguer and Prablanc [Blakemore and Donaghy, 1980; Biguer and Prablanc, 1981]. In 1987, Hine and Thorn [Hine and Thorn, 1987], presented the results of their experiments, showing the effects of near fixation on VOR during horizontal active head rotation in humans. They compared their version of the theoretical gain function that we have derived in chapter 2, with experimental data, finding that the AVOR is indeed dependent upon fixation distance; the gain increases as the viewing distance decreases. To find out which cue to distance was principally controlling the AVOR gain, they also used lenses and prisms of equivalent power, and discovered that only the second had a large effect (i.e. angular displacement and not defocusing affects AVOR gain). Their final suggestion was that AVOR gain modulation is under control of vergence and is independent of accommodation.

4.1.3 Contex modulation of VOR gain

The relationship between modulation of the VOR gain and the configuration of the oculo-motor system is further investigated by Coenen and Sejenoski [Sejnowski and Coenen, 1996]. The author's hypothesis is motivated giving a detailed description of the various anatomical structures and their interaction (shown in figure 4.8). According to these studies, otoliths and canals signals are relayed to the cerebellum and here combined with eye position and eye vergence information. The cerebellum, because of its multidimensional processing abilities, is in fact the more likely structure able to implement such complex functions of multiple signals. The two dynamic models proposed by the authors are based on the anatomical structure sketched in figure 4.8:

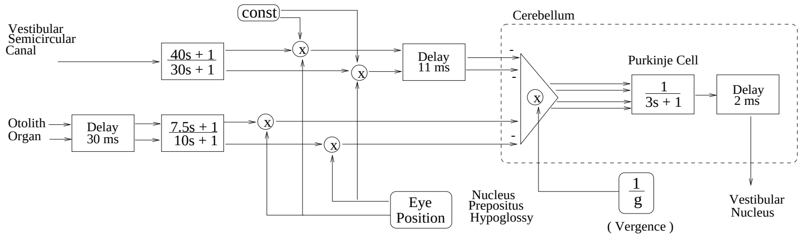


Figure 4.9: Coenen and Sejnoski model: contribution of the cerebellum to the dynamic model. Sensory inputs are combined with eye position as suggested by the kinematics of stabilization.

figure 4.9 illustrates the possible contribution of the cerebellar structure, and figure 4.10 describe the vestibular nuclei level. It is worth noting that, in both figures, the sensory signals coming from peripheral organs are combined with eye position and at the cerebellar level they are also combined with *distance to target* information (i.e. the term $\frac{1}{g}$), that is assumed to be provided by the vergence system. The predictions of their model shows to be in good agreement with VOR dynamics observed in monkeys by Snyder and King, [Snyder and King, 1992]; however the question of how the central nervous system learns the transformation hypothesized in the model still remains unknown. The cerebellum may be one site of learning for these transformation and its output may modulate in real time the VOR depending on the context; this view is compatible with other researchers results, who have found that the cerebellum is indeed required to perform correctly an otolith transformation [Angelaki and Hess, 1995].

4.2 Modeling OKR

The optokinetic reflex is a closed-loop system. During head movement, the retina provides input to cortical and subcortical pathways, where a signal "retinal slip velocity" is processed and sent to the oculo motor nuclei. Figure 4.11 shows the schematic diagram of the reflex, and its feedback nature. The first studies on the OKR were mainly conducted on animals and revealed that oculo-motor responses of foveate and afoveate animals present different characteristics. In foveate animals there is a rapid increase in the slow phase eye velocity followed by a slower rise, while in afoveate animals only the slow component is present. Moreover, it was discovered that patients with defects of the central part of the retina, causing them the lack of foveal vision, present optokinetic responses very similar to these observed in afoveate animals.

On the basis of different clinical observations, Buizza and Schmidt in 1982 proposed a model in which they made a distinction between central contribution and peripheral contribution of the retina to the reflex [Buizza and Schmid, 1982]. In their

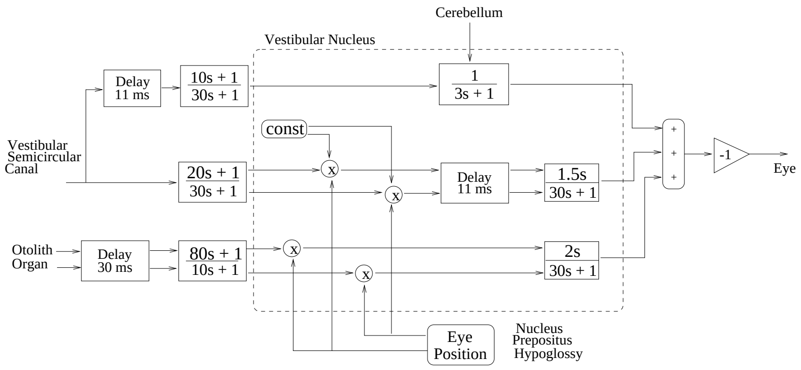


Figure 4.10: Coenen and Sejnoski model: contribution of the vestibular nucleus to the dynamical model. Three pathways in the vestibular nucleus process the canal and otolith inputs to drive the eye. The first pathway is modulated by the output of the cerebellum through a FTN (Flocculus target neuron). The second and third pathways report transient information from the input which are combined with eye position. Location of these calculation is hypothetical.

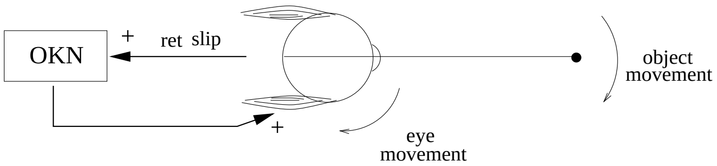


Figure 4.11: Block diagram of the optokinetic reflex.

view, the two components of the reflex are subserved by two distinct pathways, only the one providing the fast component involves the cortex and it is supported by visual receptors in the foveal part of the retina; the remaining pathway is responsible for the slow component of the response and it is primarily supported by retinal periphery. Therefore (see figure 4.12) they proposed that the signal coming from retina is split into two pathways, the cortical pathway (CP) and the subcortical pathway (SCP), the last interacting directly with VOR at the level of VN. Both stages of visual processing were identified as having non-linear characteristic, derived from saturation effects occurring in visual receptors of retinal periphery [H. Collewijn and Takahashi, 1972].

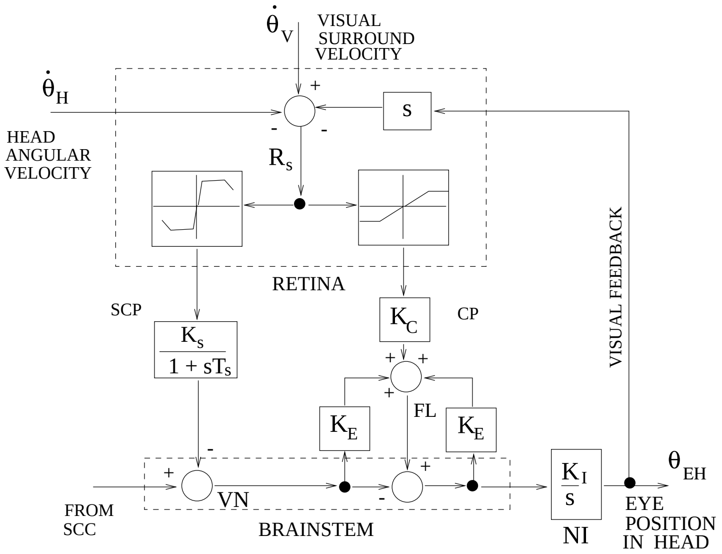


Figure 4.12: Buizza and Schmid model of OKR. Distinction is made between cortical pathway (CP) and subcortical pathway (SCP).

4.3 Modeling VOR and OKR interaction

There have been many attempt in the past to describe visual-vestibular interaction (VVI) by mathematical models. In 1970 Young proposed a model of weighted linear addition of visual and vestibular influences. The model was resolving the problem of sensory output disagreement with a switching behavior toward the most "reasonable". No emphasis on the closed loop nature of the OKR was posed by Young at that time.

It is only later, in 1977, that neuro-physiological studies focused on the association between vestibular and optokinetic responses [Waespe and Henn, 1977]. In this period Robinson and Lau formalized a much more correct description of visual stabilization mechanism. In Robinson and Lau view, visual-vestibular interaction [Robinson, 1977; C.G.Y. Lau *et al.*, 1978] is modeled by two integrated mechanisms, the open-loop response - VOR - and the closed loop response - OKR. From the viewpoint of VOR-OKR interaction, the oculo-motor control system is considered as a two inputs, one output system (see figure 4.13). The model was validated by comparing its prediction with input-output experimental data. Nonetheless, Robinson model could not explain some of the oculo-motor responses recorded in human subjects under different visual vestibular conditions (i.e. extreme stimuli condition).

A forward step in VVI modeling is due to Buizza and Schmid who considered the possibility of introducing non-linear blocks in the visual path of their previous model [A. Buizza *et al.*, 1980; Buizza and Schmid, 1982]. With such a refined model they showed that many oculo-motor responses obtained in particular experimental

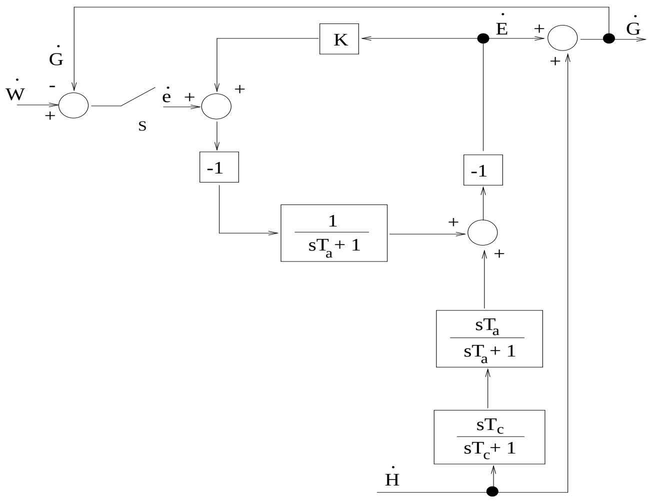


Figure 4.13: Robinson model (1977) of the vestibular and optokinetic system. The central nervous system is modeled as a positive feedback loop containing a low pass filter with constant To and loop gain K. Parameters are: $\dot{G}$ is gaze velocity, $\dot{W}$ is velocity of the world with respect to the subject and $\dot{e}$ is retinal slip velocity. The switch S represents the condition of darkness or lighted visual surround.

conditions, could be predicted correctly. They gave physiological and functional explanation for the use of non-linear blocks, accounting for saturation effects occurring at the level of the retina receptors [H. Collewijn and Takahashi, 1972].

In 1995, a much more complete description, at functional and computational level, of the overall mechanisms regulating visual-vestibular reflexes was proposed by Fred Miles [Miles, 1995]. According to Miles, peripheral end organs - canals and otoliths - generate sensory information to compensate selectively for angular and linear motion of the head. It was proved by Goldberg and Fernandez [Goldberg and Fernandez, 1975] that the canals are largely insensitive to translation and the otoliths largely insensitive to rotations in the horizontal plane. In experiments with monkeys, Miles shows that the TVOR reflex actually tunes its output with distance of object of interest (TVOR responses are linearly related to the inverse of viewing distance - see figure 4.14) and this behavior remains present even when there is no contribution from visual tracking [U. Schwarz and Miles, 1989; Schwarz and Miles, 1991].

A further observation is that each oculo-motor reflex is complemented by a visual backup mechanism that Miles and his colleagues call the OKNe - Opto Kinetic Nystagmus Early component - and the OKNd - Opto Kinetic Nystagmus Delayed component. The first generates an initial rapid rise in eye speed, the second, a gradually increasing one. This functional distinction was already addressed by Cohen et al. [T. Raphan and Matsuo, 1977] and by Buizza e Schmid, but named differently.

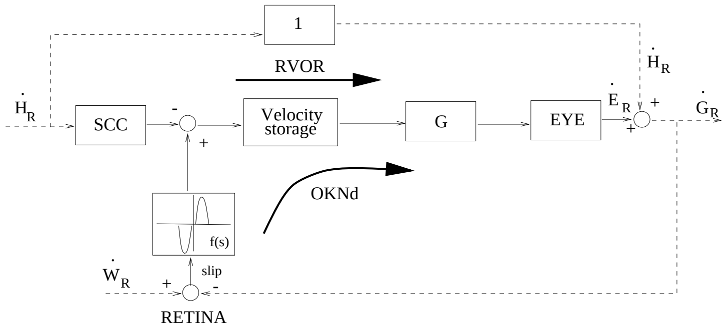


Figure 4.14: Miles et al. modeling of visual and vestibular integration: open loop RVOR and closed loop OKNd (optokinetic delayed reflex).

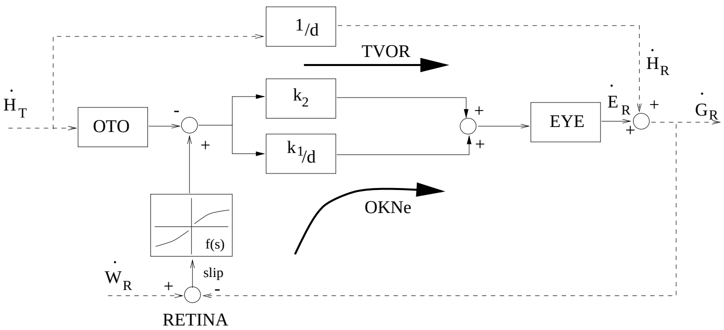


Figure 4.15: Miles et al. modeling of visual and vestibular integration: open loop TVOR and closed loop OKNe (optokinetic early reflex).

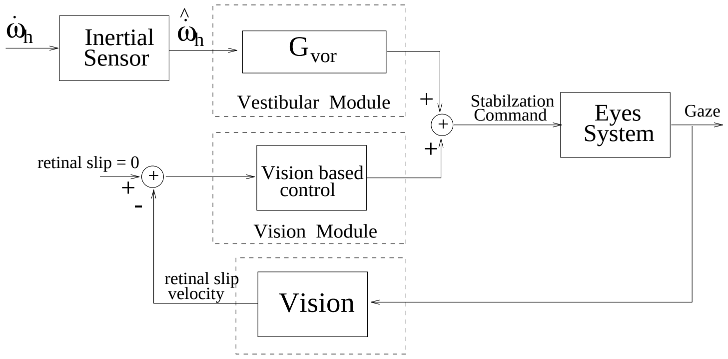


Figure 4.16: Block diagram of the control system: the inertial sensory information is processed in open loop acting in a parallel, synergistic way with to the closed loop visual feedback subsystem.

Moreover, both mechanisms work in the direction of eliminating any further residual disturbance of gaze. In particular Miles hypothesis is that OKNd evolved to compensate rotational disturbances (i.e. then coupled with RVOR) and OKNe evolved as backup to compensate for TVOR.

4.4 Modeling VVI in a monocular system

In order to explain from a robotics point of view the relative role of visual and inertial information let's consider the simple block diagram of a monocular robot system sketched in figure 4.16, where the structure of a combined visuo-inertial stabilization system is laid out.

The diagram shows that the performance of the inertial oculo-motor compensation is tuned by the gain (G_{vor}) so that, for example, a value equal to 1 would perfectly stabilize the image of points at infinity. On the other hand, the kinematics description we have devised in chapter 2 shows that a constant gain would be optimal only for constant fixation distances; on the contrary, when operating in the range of close fixations[1], rotational head movements do require continuos non-linear corrections of the amount of counter-rotational eye speed (see figure 2.2).

These observations brought us to consider the possibility to refine the model and exploit distance information to modulate the oculo-motor response appropriately; this is to say that, if the system has access to an estimate of fixation distance, its value can be used to modify the value of the RVOR gain G_{vor}. We therefore envisaged a computational model whose block diagram is shown in Figure 4.17. In this model the

[1]in the order of 20-80 cm distance

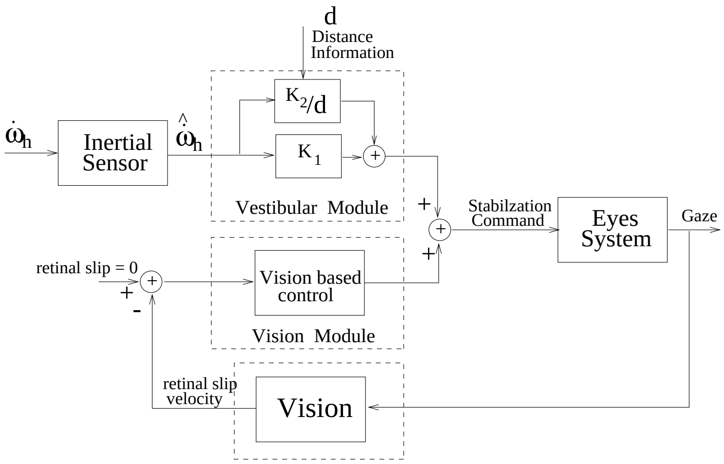


Figure 4.17: Block diagram of the control system in which distance information are considered to synthesize a better motor control strategy.

oculo-motor response is generated by the linear sum of two gain terms: a constant gain and a variable gain. The role of the variable gain term, updated on the base of actual object distance **d**, is to increase the amplitude of oculo-motor response for near fixations.

Distance information are not difficult to compute if the system is binocular and it is able to control vergence dynamically [Capurro *et al.*, 97]. This type of control keeps both cameras pointing at the object of interest, and dynamically change their orientation to maintain the object centered in the fovea, when moving in depth. It is therefore of interest at this point, to describe the basics of control of vergence, since we will refer to this mechanism in the context of our binocular modeling of stabilization.

4.5 Modeling VVI in a binocular system

The geometric arrangement of cameras in a binocular vision system is such that the knowledge of their angular orientation can lead to an estimate of the fixation distance; moreover, if the binocular robot system is able to control dynamically the steering of the two cameras, the estimate of fixation distance can be updated continuously on the basis of instantaneous cameras orientations.

In a previous work [Capurro *et al.*, 97] it was shown that control of orientation of two cameras to maintain a moving object always centered in both images can be successfully achieved using static and dynamic visual cues. We propose, at this stage, a model to integrate two oculo-motor strategies: control of vergence and stabilization.

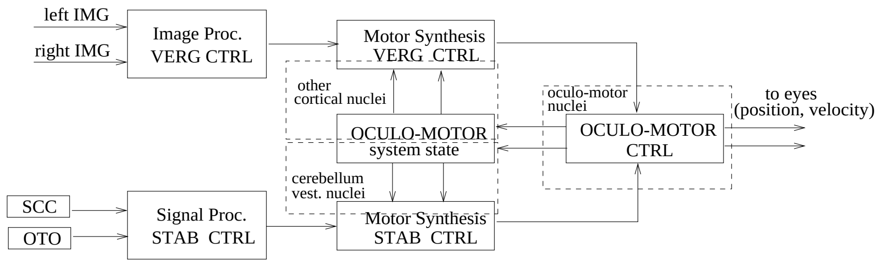


Figure 4.18: Integration of vergence and stabilization: by combining these two subsystems all the required signals to adapt AVOR and LVOR appropriately are internally available.

On the one hand, the first, provides a mechanism to obtain estimates of fixation distance; such information can be used to tune optimally the gain of the stabilization reflex; on the other, the second strategy improves the robustness of visual measures used by vergence control algorithms. In other words, integration of vergence and stabilization strategies are foreseen to support one another and to some extent are to be considered complementary; in chapter 6 the implementation of vergence control will be reviewed.

A diagram with the integration of the two subsystems is presented in figure 4.18: distance information is derived from vergence angle and is used by the stabilization subsystem to adjust the gains of AVOR and LVOR. According to this scheme, oculo-motor control for stabilization can then be adapted in real-time to different contextual situation by using information available internally to the system.

4.5.1 Fixation distance estimate

The distance of the object being fixated (see figure 4.19), can be obtained on the basis of mathematics of Viet-Muller circle. Expressed in Cartesian coordinate, the equation of these circles is:

$$(z - z_c)^2 + x^2 = R^2 \tag{4.3}$$

where $(0, 0, z_c)$ is the center and R is the radius. Figure 4.19 shows one of these circles for a given vergence angle. If we express the center and the radius according to the following equations: $z_c = \frac{i}{2\tan\delta}$ and $R = \frac{i}{2\sin\delta}$ and we use the Helmholtz parametrization[2] to describe a point (x, y, z) in space, the relation we obtain between the azimuthal disparity $\delta = \mu_l - \mu_r$ and fixation distance d is:

$$\tan\delta = \frac{ircos\mu}{d^2 - \frac{i^2}{4}} \tag{4.4}$$

[2] $x = r \cdot sin\mu$; $z = r \cdot cos\mu$, given μ, the cyclopean azimuth angle and λ, the common elevation.

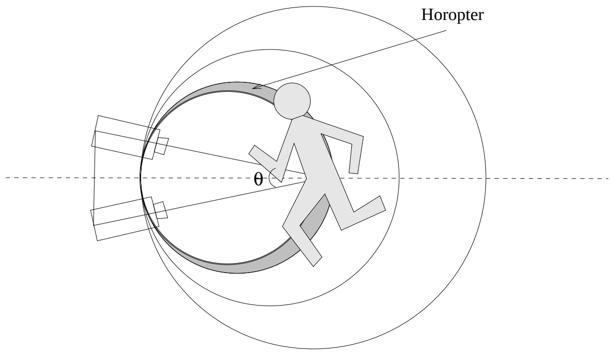


Figure 4.19: On the basis of vergence information, defining the fixation point, the Viet-Muller circle is univoquely determined.

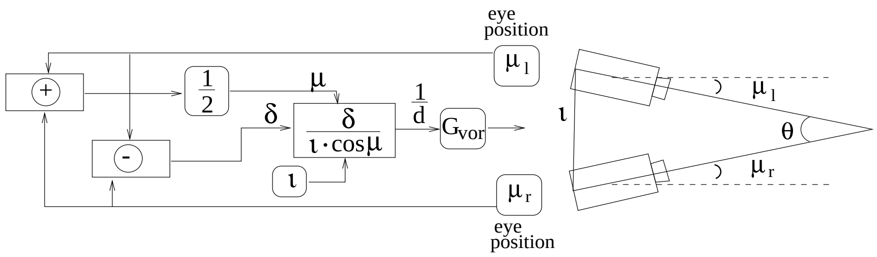


Figure 4.20: Eye position information appropriately combined leads to an estimate of distance of fixation point.

The hypothesis that interocular distance i is "small" compared to fixation distance d, further simplify the relationship 4.4 giving:

$$\frac{1}{d} = \frac{\delta}{i \cos \mu} \tag{4.5}$$

The right side of equation 4.5 shows the relevant quantities the system must know to compute the distance of fixation point: current values of azimuthal direction μ, azimuthal disparity δ and the interocular distance [C. J. Erkelens, 1997].

Figure 4.20 shows a schematic of the flow of information that need to be processed to obtain an estimate of fixation distance starting from eye position information. We stress the fact that this scheme can be easily implemented in an active vision system.

5 The Artificial Vestibular System

This chapter concerns the design of the artificial vestibular system. Three guidelines were conceived with regard to its physical realization: 1) keep the overall dimension small[1] in order to achieve a relatively simple mechanical integration of sensors assembly into a binocular vision system, 2) avoid use of sensing devices with moving parts to obtain a maintenance free structure, 3) maintain as much as possible a modular design. Modularity in particular is considered important to upgrade the whole design or parts of it: new sensor technologies and new integrated circuits are available on a day by day basis, making possible improvement of the system. Moreover, the intent of modularity was also to allow reconfiguration of the sensor array and testing of different mechanical arrangements. With regard to the second guideline, all sensors used are micro-machined solid state devices. This means that the sensing elements have no moving parts, are small and also have a low cost.

In its complete version, the multi-sensor transducer will be able to measure, three angular velocities, three linear accelerations and two angular deviations from vertical direction. In chapter 3 it has been shown that from a functional point of view, the vestibular system, provides us exactly with the above mentioned measurements. In our artificial implementation, the rotational sensors are piezo-electric type devices; the output voltage is modulated according to the instantaneous angular velocity along the longitudinal axis of the sensing element. The linear sensors, on the other hand, are based on a heat transfer principle, utilizing resistance temperature detectors as acceleration sensing elements; the change in natural convection around the sensing elements modify the output signal both in case of linear acceleration and static angular orientation. Therefore, in static conditions, the sensor acts like an inclinometer and can be used to compute orientation with respect to gravity. Both sensing elements do not contain the electronics for excitation and signal conditioning. On the one hand, the clear and direct advantage is that physical dimensions are reduced to the smallest possible; on the other hand, this approach requires the design of additional custom electronics. The following circuits were therefore realized: i) noise filtering and

[1] within a volume of 5 cm^3.

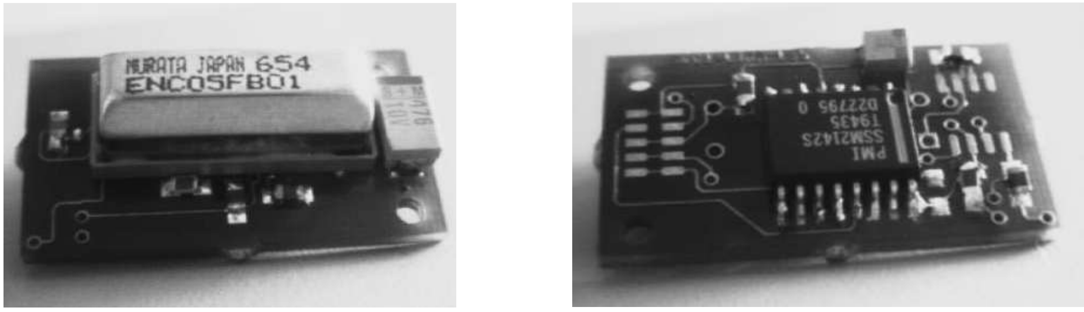


Figure 5.1: Top and bottom views of the artificial canal module: the Murata ENC-05EA sensing element, the electronics for signal conditioning and line driving are integrated on a custom designed micro-board using surface mounted technology.

bandwidth compensation, ii) thermal drift removal, iii) amplification, iv) immunity to noise for transmission over long wiring.

Two prototype assembly of single modules, have been realized. In both cases, the supports for the sensors are obtained from blocks of polyethylene. All the sensing elements are hosted internally to the blocks, respecting particular alignments. The three angular sensors are disposed in an orthogonal frame of reference so that any rotation around any given axis can be broken in three components, along three orthogonal directions; the same arrangement was conceived for the sensing elements aimed at measuring translational acceleration.

5.1 Artificial canals

In chapter 3 we have seen that during movements of the head, each semicircular canal senses head's angular velocity measuring directly the displacement of the cupola membrane. The displacement is a consequence of the force exerted on the cupola, by the endolymphatic fluid, always trying to retain its original position by virtue of its inertia.

In our "artificial canal", the sensing of angular velocity is performed using a completely different principle, known as the Coriolis "acceleration effect". A piezo-electric element is made to vibrate (for this reason also called piezo electric vibrating gyro) at constant frequency; each time the device is rotated around its longitudinal axis, the vibration amplitude is modulated by the Coriolis force experienced by the element; the amplitude modulation can be revealed by an appropriate circuit. The artificial canal is therefore composed of two parts: the sensing element and the signal conditioning electronics. While the sensing element is a commercially available product,

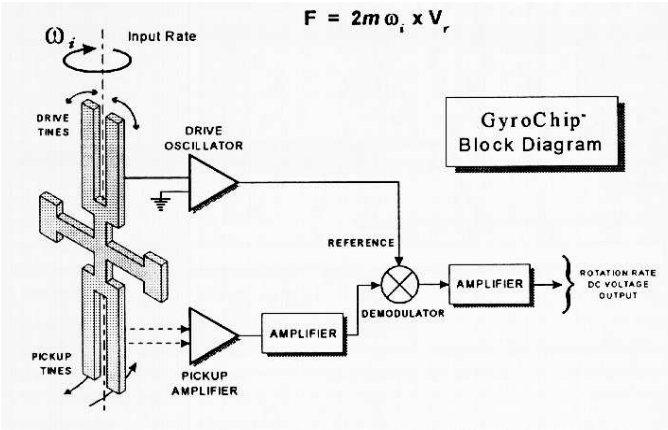


Figure 5.2: Solid state angular rate sensor: principle of operation.

(the MURATA ENC-05EA model), the electronics for signal conditioning has been entirely custom designed in our laboratory. Figure 5.1 shows the last version of the artificial canal module: the sensing element and the electronics are laid out on a double side small (1.7x3.1 cm) micro board using surface mounted components.

5.1.1 Sensing principle

Vibrating gyros, are based on a micro-miniature double-ended quartz forks. These elements can be used to sense angular velocity and are based on the well known physical phenomena expressed in the principle that "when an angular velocity is applied to an inertial body, a Coriolis force is produced in a direction perpendicular to the direction of the velocity applied". The sensor consists of a micro-miniature double-ended quartz tuning fork and supporting structure fabricated from a single wafer of mono-cristalline piezoelectric quartz as shown in figure 5.2. The drive tines, being the active portion of the sensor, are driven by an oscillator circuit at a precise amplitude, causing the tines to move toward and away from one another at a high frequency (see figure 5.2). Each tine will have a Coriolis force acting on it of

$$\vec{F} = 2m\vec{\Omega} \times \vec{V} \tag{5.1}$$

where the tine mass is m, instantaneous radial velocity is $\vec{V}$ and the input rate is $\vec{\Omega}$. The pickup tines, being the sensing portion of the sensor, respond to the oscillating torque, producing a signal at the pickup amplifier. Signal are demodulated into DC signal which is proportional to the rotation of the sensor. This sensor technology offers lightweight, inexpensive, reliable inertial sensor to directly measure rotational velocity; it opposes to traditional spinning wheel gyroscope that are heavier, consume more power and have mechanical moving parts.

5.1.2 Signal conditioning electronics

MURATA produces several version of the same sensor (i.e. the ENC-05), each with different electric and/or dynamic characteristics. In particular we have considered two versions of the ENC-05 series, with supply voltage of 3V and 5V, and bandwidth of 5 Hz extending to 11 Hz respectively. Although the S/N ratio of piezo-electric sensors is relatively high additional circuitry is always recommended to strengthen the signal. The block diagram of figure 5.3 shows the cascade of elementary stages that were developed to satisfy these requirements.

The first stage is an high pass filter. The resistor and capacitor values are chosen to obtain a pole at the frequency of about 0.3 Hz. Figure 5.4 shows the components layout designed using a CAD tool. This stage is required to eliminate the small drift effects due to temperature changes; secondly, it gives the artificial canal module an overall band pass characteristic resembling those of human's. The second stage is obtained combining an operational amplifier and an active low pass filter and enables to achieve two results: strengthen the measured signal and limit the bandwidth (also removing noise). In this case the values of the RC components of the active filter are chosen so that the cut off frequency is about 10 Hz. The third stage is a line driver and motivation for its introduction is the necessity to achieve a reliable transmission over long wiring connections. A technique known as balanced line transmission is exploited enabling to remove disturbances caused by EMI - electro magnetic interferences - occurring during transmission of signals over a long line.

The principle is explained in figure 5.5. A differential-output buffer amplifier converts a single-ended input signal to a balanced output signal pair. The signal pair is therefore transferred over the line and finally converges to a differential amplifier. Here, linear summation of the two component signal, as indicated by formula 5.2 regenerates the single ended signal and eliminate any electric noise V_D added over both lines.

$$V_{out} = (+\frac{V_{in}}{2} + V_d) - (-\frac{V_{in}}{2} + V_d) = V_{in} \tag{5.2}$$

5.2 Artificial otoliths

Head linear velocity, on the other hand, is sensed in humans by the otolith organs. Movements in any direction causes linear displacement of the macula system's receptors, the cilia, that transduce acceleration amplitude and direction into neural signals. In strict analogy with the biological principle, the class of devices most commonly used to measure linear motion, provide a direct measurement of the acceleration exploiting the displacement induced on a 'seismic' mass by inertial forces. We revise hereafter basic concepts of the class of 'seismic' accelerometers as an introduction to the principle exploited by the PEWATRON device (see figure 5.6), chosen for the realization of the artificial otolith.

5.2.1 Sensing principle

The vast majority of modern accelerometers are of pendulous type. They exploit an inertial proof mass, generally a segment of the sensing element, of known mass value. The mass is mechanically coupled with the rest of the sensor by a spring-like cantilever system. In this type of device, when the sensor is accelerated from rest, the proof mass tends to stay at rest, and the spring system is deformed. It is exactly the deformation of the spring that is transduced to become the output signal of the sensor.

On the other hand, the device that was chosen for our artificial otolith, features a heat transfer based design, utilizing temperature detectors as acceleration sensing elements. The change in natural convection around the sensor elements due to acceleration provides a stable and repeatable means to measure low $\vec{g}$ acceleration over a wide temperature range (see figure 5.7). The device is built by PEWATRON in two different versions: the single axis and double axis version. The double axis version, can measure acceleration along two orthogonal directions. According to technical data the sensing elements should respond over a frequency range of 0 to 4 Hz. In addition, this sensor has inclinometer performance through 360 deg range.

5.2.2 Signal conditioning electronics

The electronics designed for this type of sensor is described from a functional point of view in figure 5.8 and with reference to the real components layout in figure 5.9. The first stage is composed of an active low pass filter and an amplifier. The goal of this first stage is a) to remove any content in frequency not belonging to the real transduced acceleration signal, b) to strengthen the signal before stepping through the line driver stage. Moreover, since the sensor also provide a static measurement, (its DC component measure the static angular orientation), the high pass filter present in the functional diagram of the canal module, has been removed.

5.3 Canals and otoliths integration

Integration of the sensory modules is mainly a problem of mechanical arrangement. We have realized two prototypes of compound sensors: the first is composed of three angular devices, and it is therefore able to measure angular rotation around any given axis; the second, integrates a double linear accelerometer and two rotational sensors;

it is therefore able to measure two acceleration components on a plane, and rotational movements in two orthogonal directions in the plane of the linear acceleration. The drawing of the mechanical assemblies for the pure rotational sensors assembly is reproduced schematically in figure 5.10, and the picture of the real compound sensor is shown in figure 5.11; the same schematics and real images for the combined linear/angular sensors are shown in figure 5.12 and 5.13.

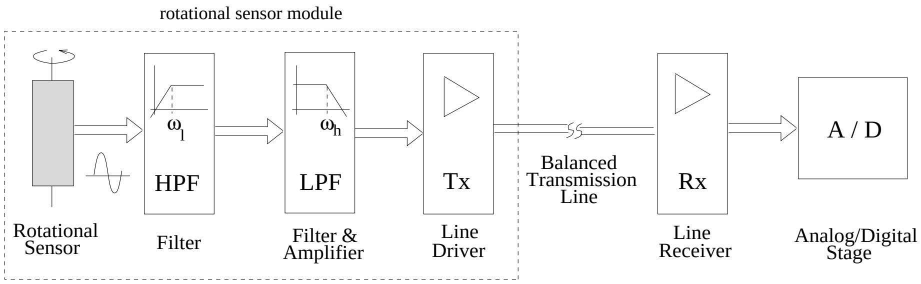


Figure 5.3: Block diagram of signal conditioning stages for the artificial canal module.

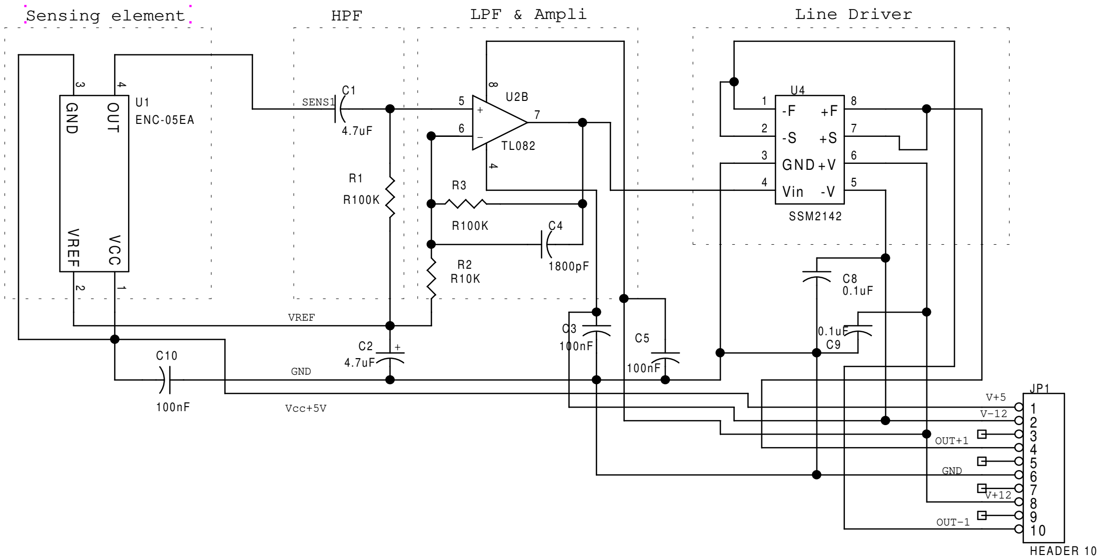


Figure 5.4: Detailed layout scheme of the artificial canal module.

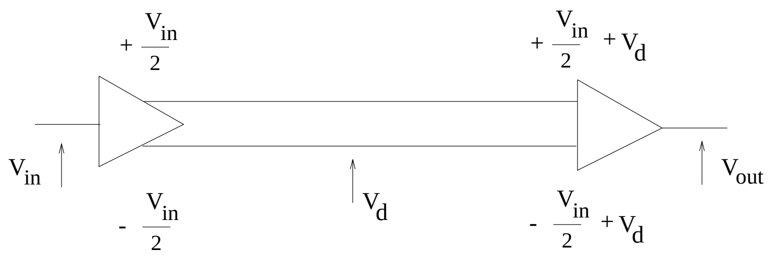


Figure 5.5: Balanced transmission technique: immunity to electro magnetic interference is obtained by means of differential information coding.

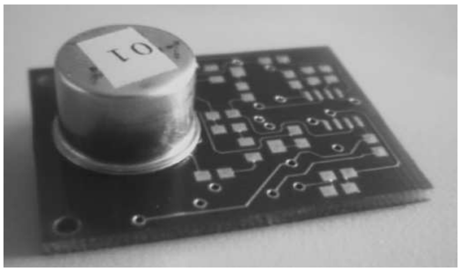


Figure 5.6: The artificial otolith module: view of sensing element placed on the micro-board. Overall dimensions are: 5x3.5x2 cm.

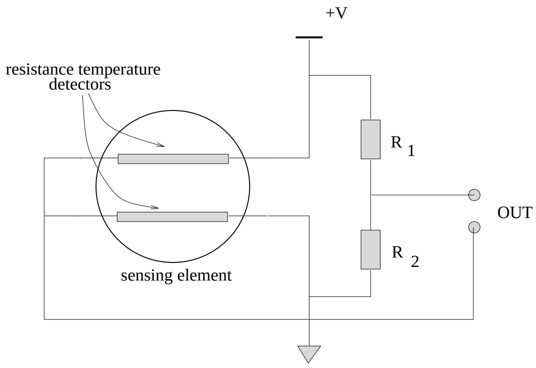


Figure 5.7: Sensing principle of Pewatron device: resistance temperature detectors are arranged in a symmetric configuration and and signal any imbalance in natural convection of air around the sensing element.

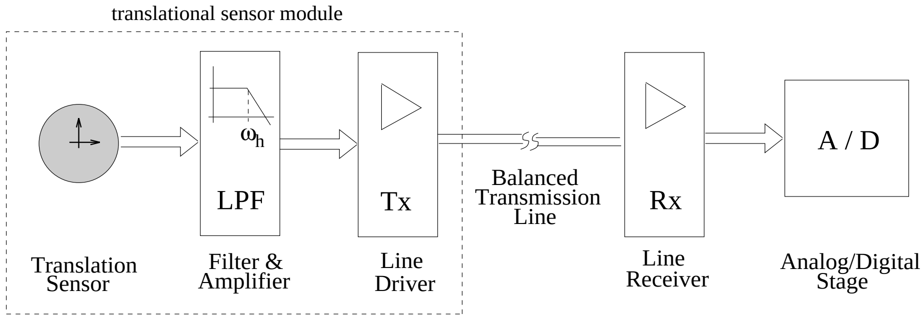


Figure 5.8: Block diagram describing the different stages in the signal conditioning electronics of the artificial otolith module.

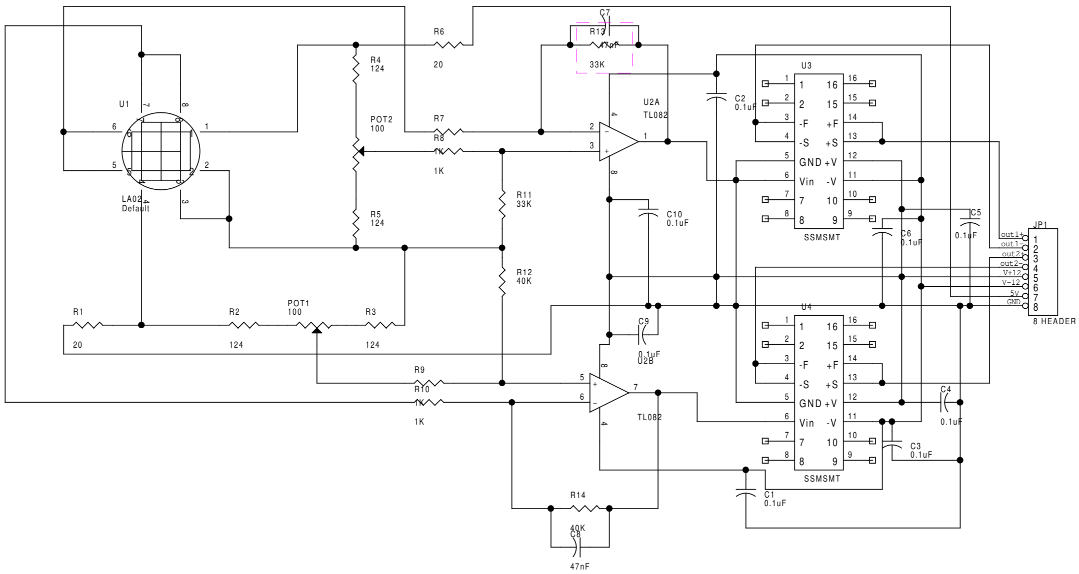


Figure 5.9: Layout scheme of the artificial otolith module.

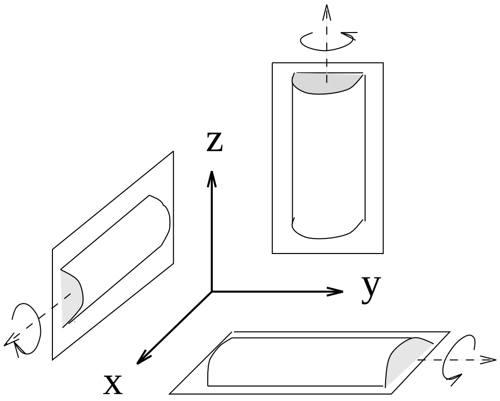


Figure 5.10: Ideal spatial arrangement of the three rotational sensory modules.

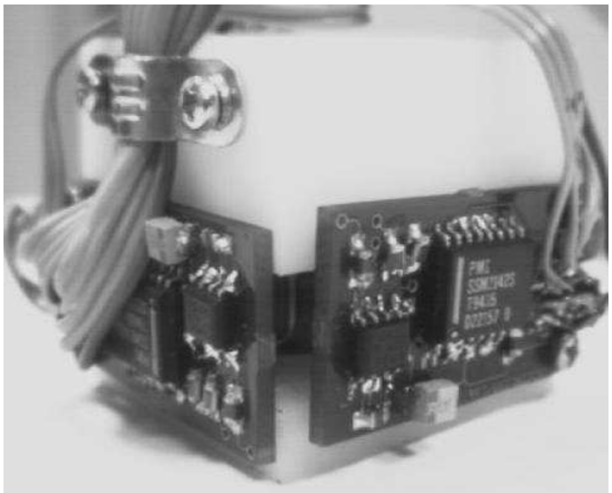


Figure 5.11: Prototype combining three canal modules. A block of polyethylene was worked out to host internally the three sensing elements; overall dimensions are 4x4x4 cm.

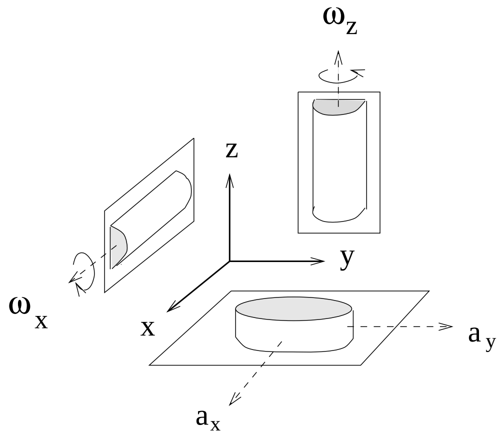


Figure 5.12: Schematic of the spatial arrangement of two rotational modules and one translational module.

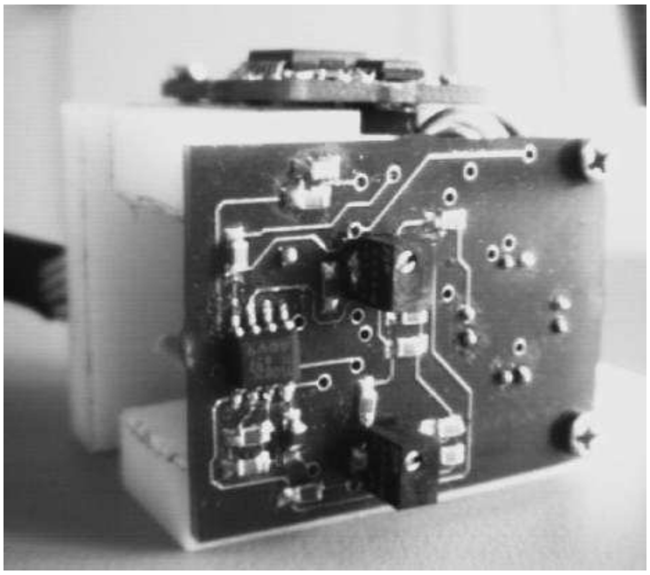


Figure 5.13: Side view of the prototype assembly combining two artificial otolith modules and two canal modules. Overall dimensions: 5x4x4 cm.

6 Experiments

In this chapter we report the experiments that were conducted to quantify the gaze stabilization behavior of the system. In particular, three experiments were envisaged, with the explicit intent of characterizing: 1) the inertial stabilization alone; 2) the visuo-inertial stabilization; 3) the performance of the system computing dynamic visual cues during visuo-inertial stabilization. Before presenting the experimental results, a description of the robotics setup, of the space variant image sensors (extended in in the appendix A.1) and of the inertial sensing device, are given. A method section was added to describe how data are collected and processed; moreover, description of the implemented control vergence are given.

6.1 Robotic setup

The robotic system apparatus is shown in figure 6.1. It is composed of:

- a 4 d.o.f binocular head [Capurro *et al.*, 1993] equipped with a pair of space-variant cameras, [Ferrari *et al.*, 1995a]; only the right camera, though, has been used in the current investigation,
- an angular rate sensor, whose driving and filtering electronics was designed and built in our laboratory,
- a Pentium 120 MHz computer and Matrox image processing boards.

In order to test the influence of all geometric parameters, the camera is positioned in a stereo arrangement with **b**=12 cm and **a**=0.5 cm. A lens with a 4.8 mm focal length is used, covering approximately a range of 50 deg. The camera is connected to a Matrox image processing board, simulating a space-variant sensor with about 2000 pixels, arranged in 32 concentric rings, each ring composed of 64 pixels. This type of sensor geometry acquires images in a “log-polar” format (see [Questa and Sandini, 1996] for a detailed description of the sensor's geometry). All visual processing is

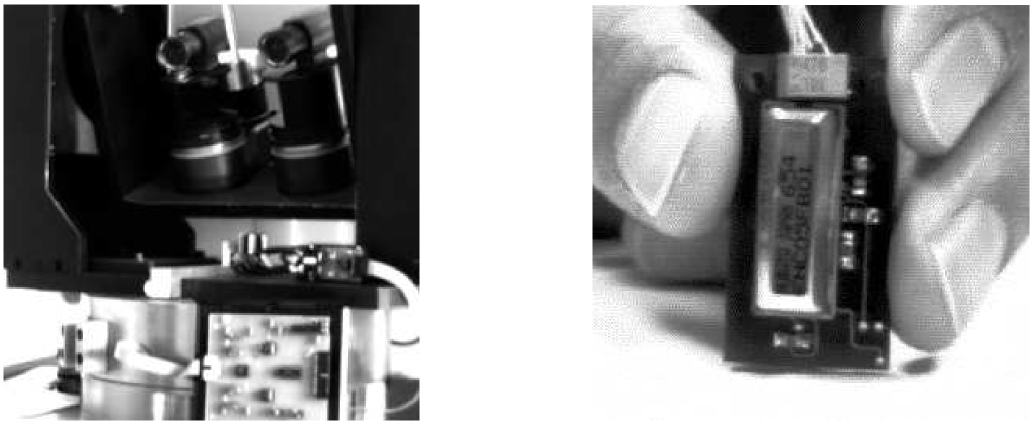


Figure 6.1: Head mount and head-mounted inertial device.

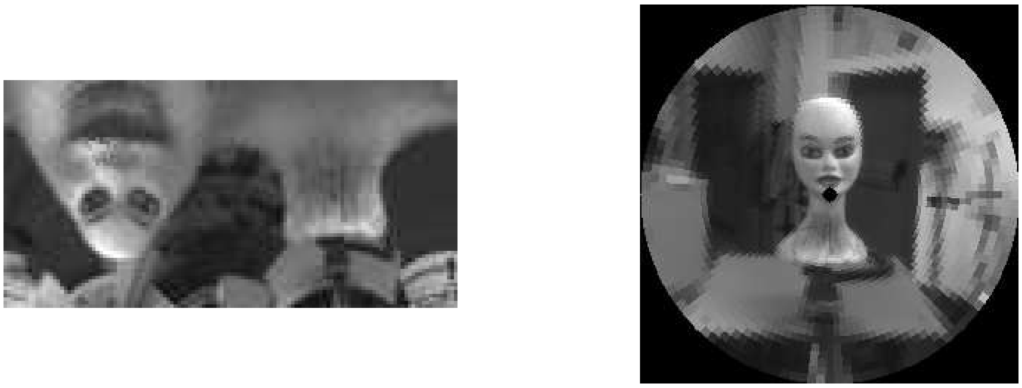


Figure 6.2: The log-polar image geometry on the left (128x64 pixels); the reconstructed Cartesian image on the right.

performed in the log-polar domain. Figure 6.2 - left shows a log-polar image (128x64 pixels), magnified to better appreciate the remapping of the foveal and peripheral areas; on the right, the corresponding Cartesian reconstructed image is displayed.

6.1.1 The inertial sensing device

The inertial sensor used is an angular rate device produced by MURATA (ENC-05EA type), which measures rotational velocity around its longitudinal axis (see figure 6.1 - right). The output of the sensor is band-pass filtered (0.3 - 5.0 Hz) and amplified using custom electronics designed at LIRA Lab. Filtering of the sensor signal gives three results: 1) eliminates the thermal drift effect, characteristic of this type of sensors, 2) determines, in response to continuos stimulation, a decay of the output

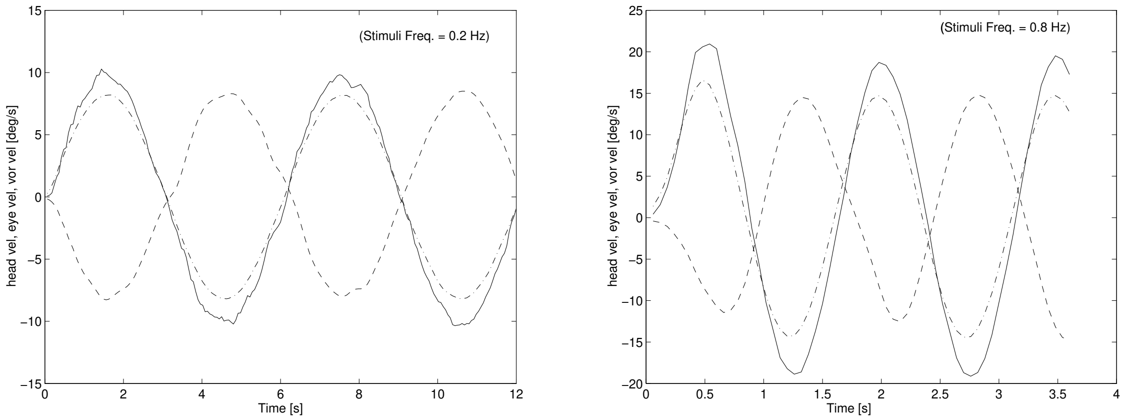


Figure 6.3: Inertial sensor output for different rotational stimuli: curves show the raw inertial signal (continuous line), the head velocity (dash-dot line) and the generated camera movement (dash line). Frequency of stimuli 0.2 Hz, peak velocity 10 deg/s(left), and 0.8 Hz, peak velocity 15 deg/s(right).

velocity signal in strict analogy with the biological system's response, 3) filter out effects of mechanical vibrations and electronic noise present in our robotic setup. On the other hand, according to technical data provided by MURATA, the sensor itself is able to provide measurement in a range of motion extending up to 20 Hz. The sampling is performed with a 12 bits A/D converter; the maximum angular velocities measurable by the sensor are, approximately, -90 deg/s and +90 deg/s.

Figure 6.3 shows, in continuous line, the sensor output for two different motion stimuli (different frequencies of head oscillation); the sensor output is used here directly as a control signal to generate camera movements with a constant G_{vor} gain of 0.95.

6.2 Methods

6.2.1 Generation of inertial stimulus

In order to measure the system performance, a rotational stimulation of the inertial sensor is produced, by imposing on the head azimuthal axis (i.e. head pan axis) a measurable rotational motion profile. Frequency and amplitude of the oscillatory motion can be varied within the hardware admissible range. Stabilization of gaze is obtained controlling one degree of freedom: the pan of the right camera. The camera tilt angle is not controlled during the current experiments. The motion profiles of inertial stimulus (i.e. head pan axis) and the eye response are measured reading encoder position.

All experiments are performed in real-time with a control period of about 80 msec[1]. During the experiments the following parameters are stored to estimate the overall performance: i) position angle of the inertial stimulus; ii) output of the inertial sensor; iii) u_0 component of optic flow; iv) values of the image stabilization index; v) position angle of the controlled camera. Head and eye velocities are computed off-line on the basis of the sampled angular position using a 5-points approximation of derivative.

6.2.2 Residual Ocular Velocity

The contribution of the inertial sensory pathway to oculo-motor compensation can be better understood by examining figure 6.4. The curves show the eye velocity required to maintain a stable fixation for different constant values of G_{vor} at various distances **d**. The angular velocity of the head considered in this example is $\dot{\omega}_h = 10\ deg/s$. It is worth noting that with a fixation point at 1 m a visually-driven compensation of 12 deg/s is required without inertial stabilization ($G_{vor} = 0$); this value drops to about 2 deg/s for G_{vor}=1. From the visual processing point of view it is precisely this velocity (which we call *Residual Ocular Velocity ROV*) that needs to be measured from image velocity information; the higher the ROV, the more computationally intensive is the required visual processing.

In order to know the amount of visual processing required to compensate the ROV, an estimate of image velocity as function of the geometric parameters of the head-eye system and fixation point distance has to be derived. The equations describing the image velocity due to relative motion between the eye and the scene, as function of eye's motion parameters T_x, T_y, T_z, W_x, W_y, W_z are [Sundareswaran, 1991]:

$$u(x,y) = f_x \left[\frac{\frac{x}{f_x}T_z - T_x}{Z(x,y)} + W_x \frac{xy}{f_x f_y} - W_y(1 + \frac{x^2}{{f^2}_x}) + W_z \frac{y}{f_y} \right] , \tag{6.1}$$

$$v(x,y) = f_y \left[\frac{\frac{y}{f_y}T_z - T_y}{Z(x,y)} + W_x(1 + \frac{y^2}{{f^2}_y}) - W_x \frac{xy}{f_x f_y} - W_z \frac{x}{f_x} \right] . \tag{6.2}$$

To carry out this analysis the horizontal component $u(x,y)$ can be simplified, neglecting the translational and rotational terms T_y, W_x, W_z and considered at the center of the image (i.e. $(x,y) = (0,0)$). These assumptions are based on the hypothesis

[1] the computer architecture used at the time data were collected did not support multi-tasking; this constrained the latencies of the visual and inertial information to a common value.

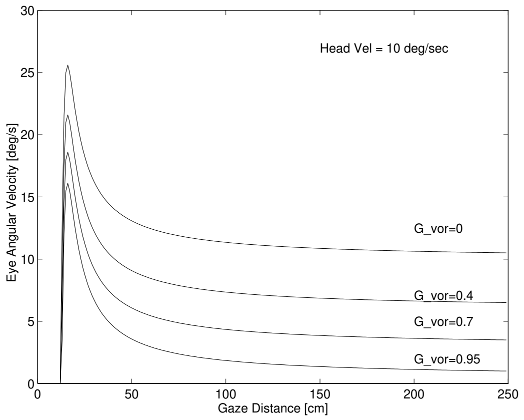


Figure 6.4: Angular velocity of the left eye for different G_{vor} values. The baseline $\mathbf{b}$ is set to 6 cm, $\mathbf{a}$ is equal to 12 cm and $\dot{\omega}_h$ is equal to 10 deg/s.

of considering rotations around the Y axis (i.e. vertical) and translations on the ZX plane only. This yields to:

$$u(0,0) = f_x \left[\frac{-T_x}{Z(0,0)} - W_y \right] . \tag{6.3}$$

If the absolute angular velocity of the eye W_y is expressed in terms of the eye velocity relative to the head $\dot{\omega}_e$= -G_{vor} $\dot{\omega}_h$ (where $G_{vor}$is constant and $\dot{\omega}_h$a given head velocity),

$$W_y = (1 - G_{vor})\ \dot{\omega}_h \tag{6.4}$$

and T_x is expressed in the eye reference system (i.e. considering the relative orientation ω_e of the eye with respect to the head coordinate system)

$$T_x = \left(\frac{\mathbf{b}}{2}\ \cos\omega_e - \mathbf{a}\sin\omega_e \right)\ \dot{\omega}_h\ , \tag{6.5}$$

equation (6.3) becomes:

$$u(0,0) = \left[-\frac{\frac{\mathbf{b}}{2}\cos\omega_e - \mathbf{a}\sin\omega_e}{Z(0,0)} - (1 - G_{vor}) \right]\ \dot{\omega}_h, \tag{6.6}$$

where $Z(0,0)$ is the actual distance of gaze point $\mathbf{P}$ from eye position $\mathbf{E}$. This relationship can be used to estimate the average image translation u_0 measured in the "center" of the image plane for different values of G_{vor} gain, in response to different head velocities $\dot{\omega}_h$. In the singular case of $G_{vor} = 0$, the equation 6.6 provides an estimate of the computational requirements of a system using a "pure" visual stabilization approach.

6.2.3 Measure of stabilization performance

In order to asses quantitatively the performance of the stabilization loop for many different experimental conditions (e.g. different G_{vor}values, different distances), we compute the Image Stabilization Index (ISI) defined in the following way: $ISI_i = 1 - NC_i$, where

$$NC_i = \frac{\sum\limits_{\eta,\xi}(I_i(\eta,\xi) - \mu_i) \cdot (I_{i-1}(\eta,\xi) - \mu_{i-1})}{\sqrt{\sum\limits_{\eta,\xi}(I_i(\eta,\xi) - \mu_i)^2 \cdot \sum\limits_{\eta,\xi}(I_{i-1}(\eta,\xi) - \mu_{i-1})^2}} \tag{6.7}$$

is a normalized correlation index between two subsequent log-polar frames (I_i and I_{i-1}); μ_i, μ_{i-1} are their mean values. Better performance is therefore mapped to lower values of ISI. It is worth noting that this visual measurement is indeed the simplest direct way to quantify stabilization performance correctly, particularly when dealing with active vision systems and multiple control loops acting in real-time. In these cases estimation of delays is sometimes difficult and, therefore, an analysis based only on angular positions may give misleading results.

6.2.4 Measure of image velocity: range and accuracy

For the purpose of our measurements a first-order approximation (i.e. affine model) of the optic flow is sufficient and easier to compute [Koenderink and van Doorn, 1991]. In particular we are interested in measuring the horizontal component in the center of the image; in fact, this estimate is directly proportional to the amount of ROV that can be canceled out by appropriate camera rotation (eye movement). To compute the values of the affine parameters from log-polar images, the same estimation method already used in the context of dynamic vergence experiments [Capurro *et al.*, 97] is adopted. It is in appendix A.2, that we revisit the affine model of the optic flow and how image velocity information is extracted in the log-polar domain. It is important, at this point, to know quantitatively the range and the accuracy of the estimate of u_0 component of the optic flow as measured by our algorithm. We derived this information experimentally by repetitive measurements: a camera fixed on top of a controllable slide is repetitively translated backward and forward at constant velocities while the u_0 estimates are collected. The camera is moving in front of a textured flat surface, 245 cm distant, along a front-parallel trajectory. At the end of each back-and-forth motion sequence, the velocity of the slide is increased. The range of translation velocities is between 1 cm/s and more than 20 cm/s . Figure 6.5 shows the u_0 component (left-side) and the image stabilization index (ISI) (right-side) as function of translation velocity. The mean value and the standard deviation are plotted for each experiment. The data show a good behavior of the u_0 estimate for translation velocities up to 17 cm/s. For higher translation velocities (more than

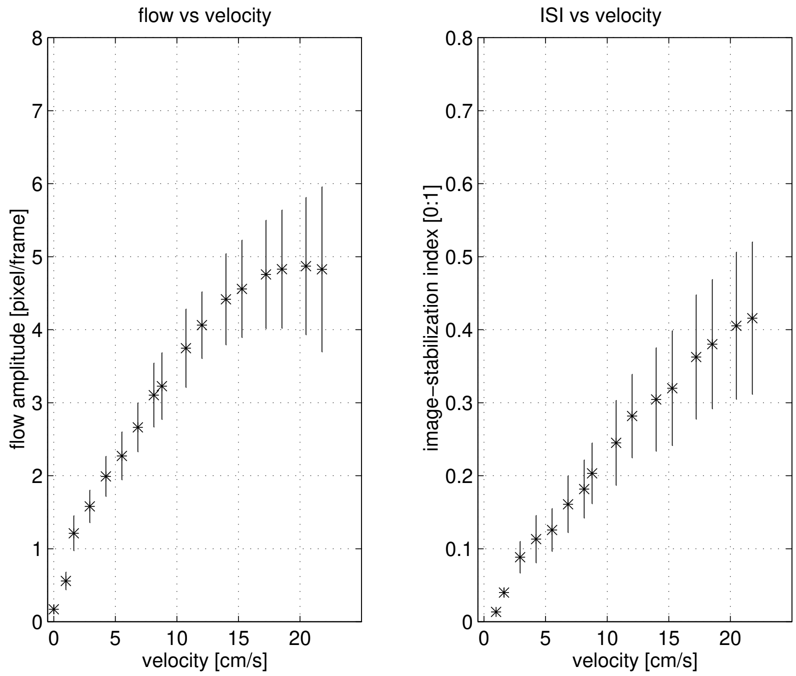


Figure 6.5: Range and accuracy of the optic flow horizontal component u_0 as measured by our algorithm: saturation effect (left), and image stabilization index (right).

20 cm/s) the u_0 estimate saturates. These recordings, once again, emphasize in a quantitative way that, for a given amount of computational resources[2], visual information alone can be used reliably only within the "good range" of measurements, that is, up to the point were a "saturation effect" starts corrupting the accuracy and robustness of the visual measuring process. In the discussion section, this point is further elaborated. We will numerically compare gaze stabilization requirements, in terms of ROV (i.e. residual image velocity), that must be robustly estimated, to generate optimal stabilization, either using visual information, or conversely, using inertial information.

6.2.5 Vergence control

Vergence movements enable binocular fixation of points in visual space at different distances from the observer; our robotic implementation relies on visual information and active movements of cameras. Two visual quantities are used to drive camera motion: the temporal evolution of a measure of binocular fusion and the change in scale that arise on the image plane as a consequence of the relative motion between an object and the observer, measured, in our case, by the divergence component of optic flow.

The degree of fusion of the stereo pair is computed on the basis of a cross correlative measure (the so called *fusion index*) applied between the left and right images; the change in scale, computed on the *fused image*[3], is derived from the estimate of the divergence component of an affine model of the optic flow. Under some assumption, divergence is proportional to the change in depth of the fixated object. In principle the use of a global correlation index between the images of the stereo pair is not sufficient to build a motor reflex similar to what in biology is known as *compulsion for vergence* [Ogle, 1964] because global correlation lacks any information regarding the motion direction in depth. However, with reference to figure 6.6 depicting a general experimental situation for vergence control (left), if active movements of the cameras are introduced, the correct steering direction can be obtained from the derivative of the fusion index with respect to the vergence angle. The plots shown in figure 6.6 (right) represent the value of the fusion index with respect to the vergence angle, during a passive vergence motion of the two cameras and for different angular velocities. In spite of the coarseness of the plotted curves for the fast velocities (the processing time is constant in all experiments), the minimum can be easily identified and it can be observed that the derivative of the fusion index is low when the cameras are far from the correct vergence configuration (see figure 6.6-right between 4 and 5

[2] the computational resources used in our experimental setup consist of: a Pentium 120 MHz, Matrox Images frame grabber/pre-processor, parallel port interface for sensor reading - after digital sampling.

[3] the fused image is obtained by linearly adding left and right images in foveal areas where high level of matching occurs.

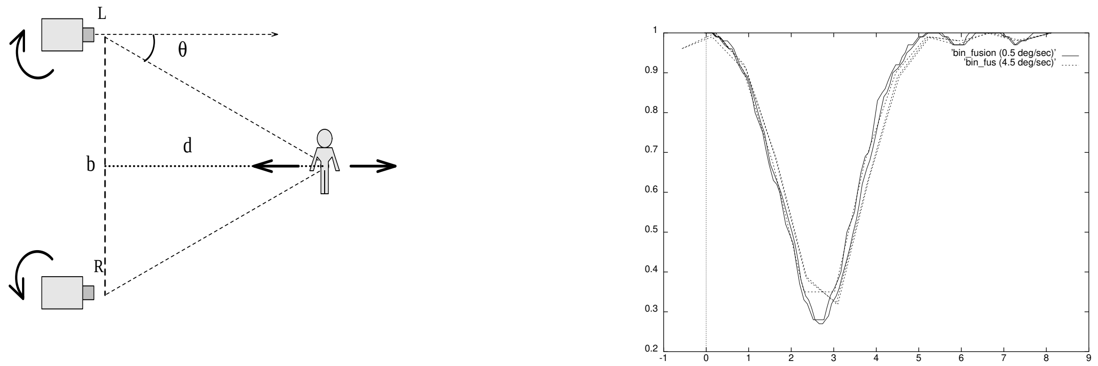


Figure 6.6: (left side) Symmetric camera rotations and relative movements between target and vision system, give rise to V-shaped binocular fusion profiles (right side). The curves on the right are obtained by verging/diverging cameras symmetrically at constant speed while a person is standing in front of the vision system (left side). The two plots correspond to different camera angular velocities, respectively 0.5 deg/sec and 4.5 deg/sec.

degrees) while it is higher when the camera approaches the correct vergence angle (i.e. between 2 and 3 degrees). For this reason, the derivative term appears in the denominator of the velocity signal used to control the motion of the camera in the following way:

$$c_{v\_corr} = -\frac{C(\theta)}{\nabla C(\theta)}$$

The differential correlation term represents the steepness of the profile plotted in figure 6.6: a positive value indicates that the system is working on the right side of the curve, a negative value, identifies the left side of it. On the basis of this information the system is then capable of choosing a *convergent* or a *divergent* motion and drive the cameras towards the minimum of the correlation profile. The term $C(\theta)$ at the numerator of the control law has been introduced so that the control signal c_{v_corr} goes to zero smoothly as binocular single vision is reached, reducing the effects of noise introduced by the derivative operator.

Once binocular single vision is reached by fusing left and right log-polar images, the camera motion gets down to zero (i.e. $\Delta\theta = 0$) and the previous control law becomes ineffective. When the cameras are steadily verging the system must introduce a second control strategy. The previous control law alone cannot account for

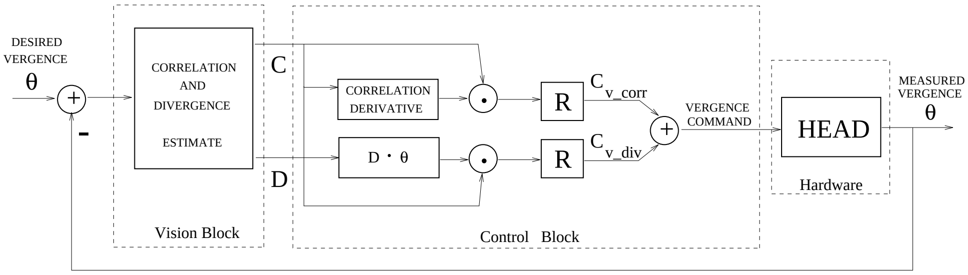


Figure 6.7: Vergence control scheme. Correlation, correlation derivative with respect to vergence angle and divergence estimate are combined together and input to standard regulators (e.g. PID) represented by R blocks.

the kind of dynamic behavior we want to achieve. Consider, for example, the case of a fixated object which suddenly starts moving in depth. By only using the fusion index no information would be available regarding the initial direction of the vergence command (convergence vs. divergence) and the only sensible solution would be to chose randomly and verify if the selected direction is causing the fusion index to decrease or increase. Without explicitly using disparity information and/or accommodation the only way to compute the correct initial vergence direction is by measuring the expansion component of the affine motion field. This measure takes into account any change in distance between the fixation point and the eyes of the observer. The divergence component of the optical flow (which is somehow simplified in the log-polar plane), allows the estimation of such motion information. Consider now Fig. 6.6 (left) and the the approximate relation $\frac{\Delta\theta}{\theta} \approx -\frac{\Delta d}{d}$. By dividing both terms by Δt and substituting the divergence expression given in terms of velocity and distance to target $D = 2\frac{\dot{d}}{d}$ the following can be derived $\dot{\theta} \approx -D(t) \cdot \frac{\theta}{2}$. Taking into consideration both the fusion index and divergence a second control strategy can be synthesized as follows:

$$c_{v\_div} = -C(\theta) \cdot D(t) \cdot \frac{\theta}{2}$$

where the term *D(t)* represents the divergence measurement. Positive correlation index, introduced for the same reasons mentioned before, coupled to dual-signed divergence, gives rise to a signed velocity command used to drive the vergence controller. Vergence control that uses correlation measurements coupled with expansion/contraction stimuli implementing a cross-talk signal structure (see figure 6.7), results in a very robust and efficient motor reflex behavior.

6.3 Inertial stabilization experiment

To stress the role played by the inertial sensor contribution, we started each experiment, with the parameters of the rotational stimulus (amplitude and frequency) set to values sufficient to induce, on the image plane, a motion field close to saturation threshold of the visual processing system (compare the residual optic flow of figure 6.9 - left for G_{vor}= 0 and range/accuracy data of figure 6.5). It is worth noting that, in such a "relatively high" dynamic conditions (10 deg/samplitude, 0.2 Hzfrequency), a performance comparison with pure visual stabilization cannot be performed correctly.

In the first experiment stabilization is performed using inertial information alone. The ROV measured in the center of the image plane, reflects dependence on fixation distance $\mathbf{d}$ (specially in 0-250 cm range) and inertial compensation gain G_{vor}. Data are collected for distances $\mathbf{d}$ of 130 cm, 180 cm, 240 cm and G_{vor} at 1., .7, .5, .3 and 0 (no compensation).

A sample output of the data recorded during stabilization is shown in figure 6.8. With respect to these (and following) figures we would like to point out that the profile of the head velocity is far from sinusoidal. This is a consequence of the way we generated the velocity profile of the inertial stimulation [4].

Figure 6.8 (top-left) shows the curves of head velocity, inertial measure and eye velocity for a value of G_{vor}=0.7. The "*" symbol superimposed to the head velocity profile, identifies the relative maxima of head velocity. Figure 6.8 (upper-right) shows the computed horizontal component of the optical flow u_0. As it can be seen here, the maximum value of the computed optical flow is already close to saturation level of the algorithm (compare with results shown in figure 6.5). In figure 6.8 (bottom), the stabilization index is shown for the same experiment. The "*" symbols here indicate the values of the ISI corresponding to the maxima of the head velocity profile. These plots are given here to show qualitatively the performance of the inertial stabilization and the performance of the optic flow computation.

In order to quantify these parameters for different values of $\mathbf{d}$ and G_{vor} the same data are recorded in other experimental sessions and representative numerical values are computed off-line on the acquired measures. This was done by identifying the maxima of the velocity profiles of the oscillatory motions applied to the camera (i.e. the maxima of the velocity profiles used to stimulate the inertial sensor). These points are indicated in figure 6.8 by "*" symbols and are used as time references to extract from the optical flow and ISI data the values to be used as representative data. In all experiments, at least 10 maxima are identified in the head velocity profile, and the

[4] the servoing of the main head axis for the generation of sinusoidal trajectories was obtained using an external controller. This solution avoided any interference in the stabilization control loop; unfortunately due to some hardware compatibility problem, an optimal tuning of servoing parameters, and consequently, optimal generation of sinusoidal trajectories, was not possible.

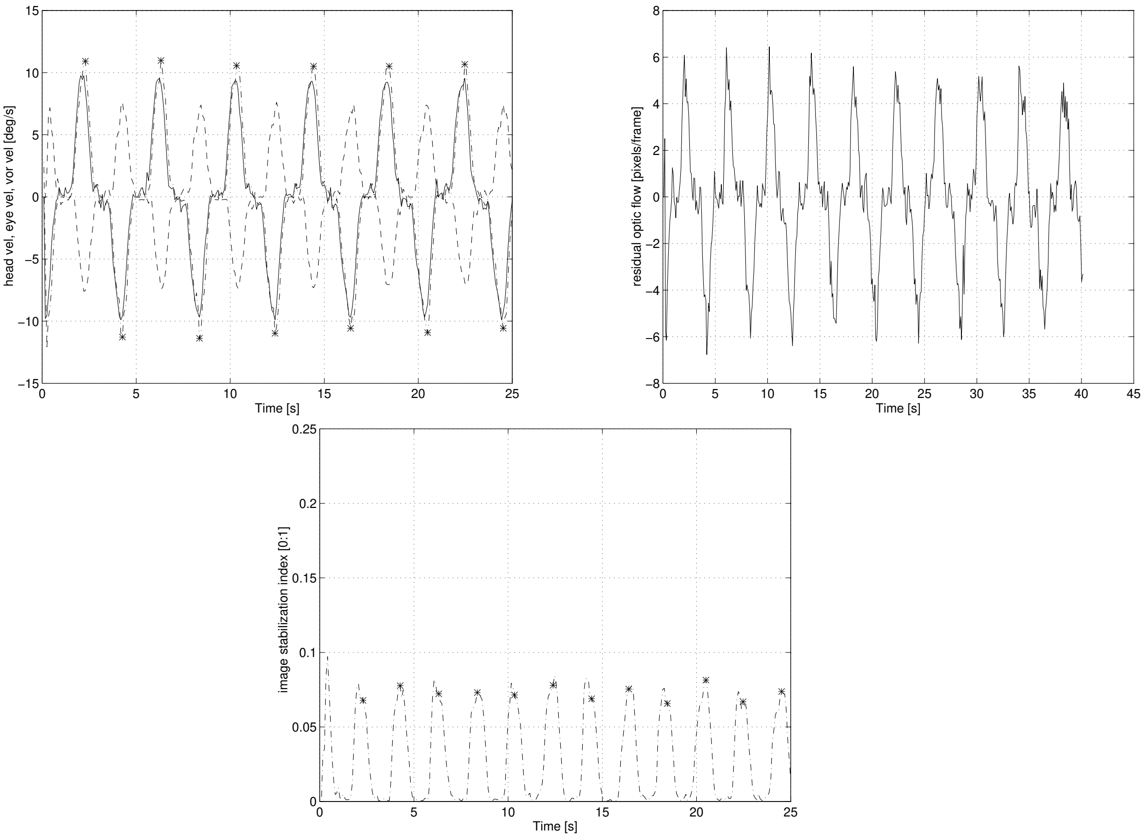


Figure 6.8: Sample data during one experimental session: (top) head velocity (dash-dot), inertial sensor output (continuous) and eye velocity (dash). "*" symbols on the head curve identify the maxima of the head velocity profile and are used as time references for optic flow and correlation measurements. In this particular case the value of G_{vor} is equal to 0.7 and the distance of the fixation point is 240 cm. (bottom) Image stabilization index (ISI) recorded during the same experiment.

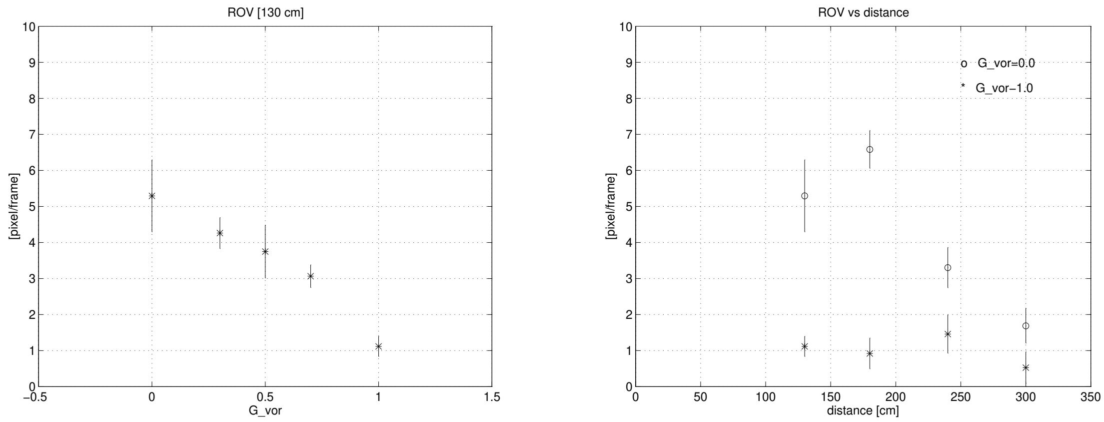


Figure 6.9: Residual optic flow measured on the image plane. (left) Fixation distance is 130 cm: a decrease of 0.5 in VOR gain, increments by a factor of 3 the ROV. (right) ROV for different distances when two different G_{vor} gains are used.

mean and standard deviation of the optical flow and correlation at this time instants computed.

The variation of ROV as a function of G_{vor} is shown in figure 6.9 (left) for a fixation distance of 130 cm. The relationship between G_{vor} and ROV is evident and it is worth noting that, without inertial stabilization (i.e. for $G_{vor} = 0$), even at this relatively long distance ROV approaches the saturation level. If the object gets closer and no inertial loop is present, stabilization based only on vision would be impossible. This is clear in figure 6.9 (right) where the value of ROV for various distances and two different values of G_{vor} is shown. In fact, while the values obtained for G_{vor}=1 (tagged with a '*' symbol) are well maintained within the range of our optical flow estimation, the values at 180 cm and 130 cm for G_{vor}=0 (tagged with an 'o' symbol) begin to saturate, giving a wrong estimate. The consequences of this can be seen in figure 6.10 where the ISI is plotted for different values of G_{vor} and for two different distances of the fixation point (remember that, according to our measure, a high value of the ISI means a bad stabilization).

6.4 Visuo-inertial stabilization experiment

In the second set of experiments the integration of inertial and visual information was implemented. A constant value of G_{vor} is chosen (close to the human VOR gain in darkness) and the residual "retinal-slip" is compensated using visual information. An improved performance is indicated by the image stabilization index (ISI) over different fixation distances. The u_0 component of the affine flow was used to synthesize an

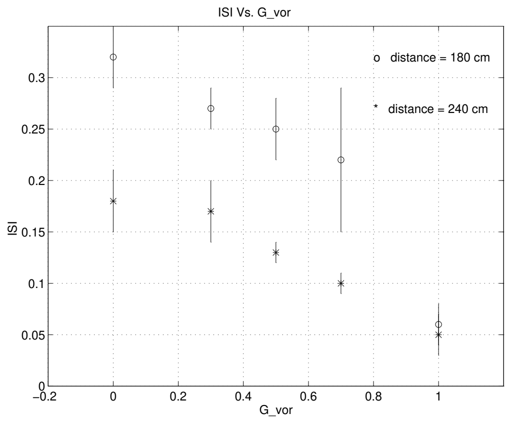


Figure 6.10: Image stabilization index: better stabilization is mapped to lower index values. Stabilization behavior is shown for two different fixation distances (180 symbol 'o', 240 symbol '*').

oculo-motor command that was linearly added to the inertial information. At the present stage, the control strategy is very simple and does not take into consideration the possibility of different delays in the two control loops (i.e. visual loop and inertial loop). However, even with this simple control strategy, the advantages obtained using both, visual and inertial information, are evident. The results of these experiments are shown in figure 6.11. Two sets of residual optic flow and the ISI data are plotted, each set corresponding to a constant G_{vor} value (i.e. $G_{vor} = 0.7$ is represented by 'o' symbol, and $G_{vor} = 1.0$ data represented by '*' symbol). Data within each set are obtained increasing the amount of visual information used to synthesize the final stabilization command (i.e. increasing the visual gain). Performance in stabilization, measured by the image stabilization index (ISI), improves for both values of G_{vor}. As the use of the visual component (i.e u_0) in the control loop increases (i.e. by increasing the vision gain), the resulting ISI becomes smaller.

In figure 6.11 (left) the residual optic flow (ROV) for visual gain = 0.0 (i.e. only inertial information is used in the control loop) is estimated in 2.3 pixels/frame for G_{vor}=1.0 and 4.5 pixels/frame for G_{vor}=0.7. This values reduce almost to zero pixels/frame and to less than 2.7 pixels/frame when visual information is introduced with a gain equal to 1.5. The resulting improvement in stabilization is also evident in the ISI plots (see figure 6.11 - right), where the ISI linearly decreases, thus indicating better overall image stability. Figure 6.12 shows in binocular frames, the simultaneous left camera view (non-stabilized) and the right camera view (stabilized) of the vision system, during a sinusoidal oscillatory head motion[5].

[5]these images are recorded directly from the two cameras while the processing is performed on log-polar images

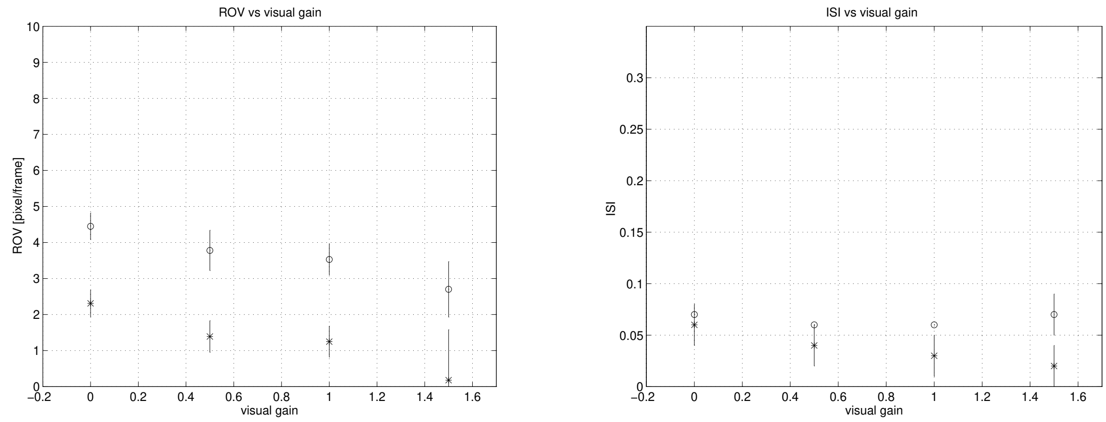


Figure 6.11: ROV and ISI of two sets of data. G_{vor} is constant within each set, while visual gain is increased; 'o' symbol for $G_{vor} = 0.7$; '*' symbol for $G_{vor} = 1$.

6.5 Dynamic visual measurement during stabilization

The visuo-inertial stabilization strategy is tested in relation to a dynamic visual measurement task, by fixating an approaching object while the head is oscillated sinusoidally at a given amplitude and frequency (see figure 6.13). In this situation the expansion/contraction pattern produced by object motion on the image plane is estimated by computing divergence (see appendix A.2 for details on divergence measurement), a positive divergence indicating that the object is approaching, a negative divergence indicating the object is moving farther away. Figure 6.14 (left) shows the estimate of divergence during visuo-inertial stabilization. The profile of divergence clearly alternates between positive (object approaching) and negative values (object moving away). It is in these operative conditions, that the estimated divergence can be used successfully in a control loop (e.g. to control vergence dynamically [Capurro *et al.*, 97]).

On the contrary, figure 6.14 (right) shows divergence measured in non-stabilized gaze condition. The clear alternation between positive and negative "behavior" disappears and use of this visual cue as a parameter to control camera movements will be inappropriate.

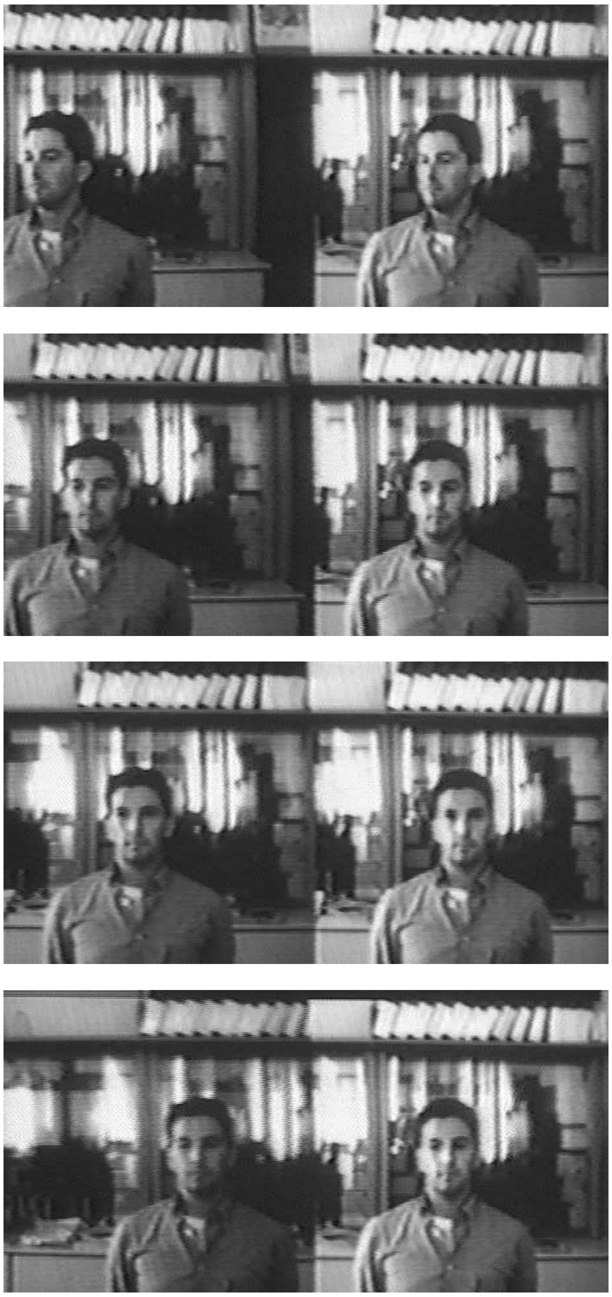


Figure 6.12: Binocular frames during stabilization experiment: non stabilized camera (left), and stabilized camera (right).

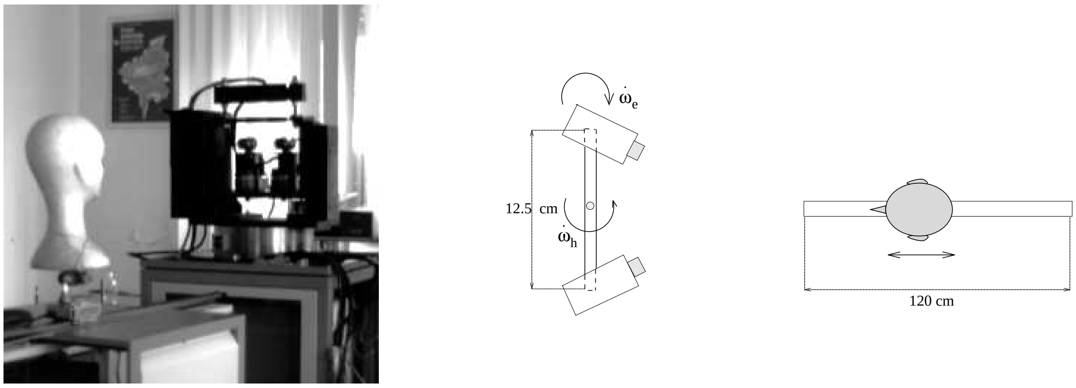


Figure 6.13: Setup used for estimating divergence during oscillatory head motion.

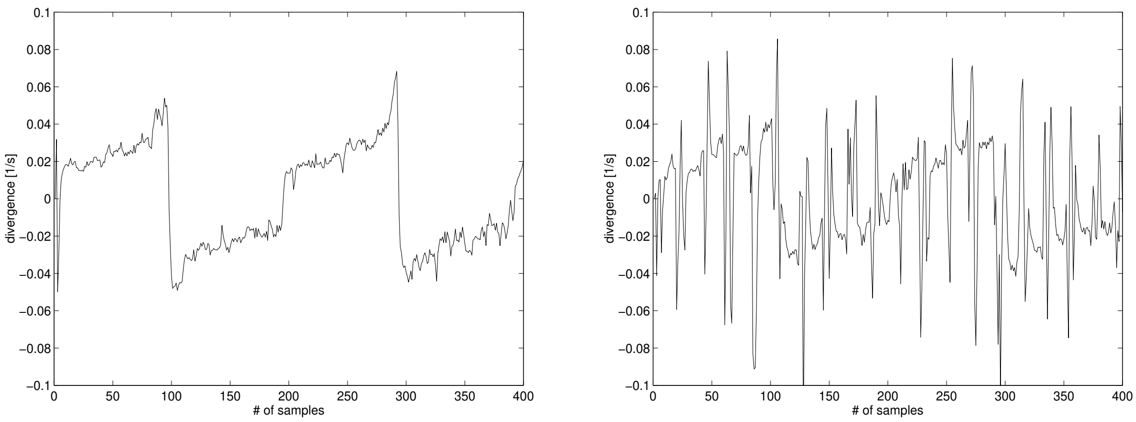


Figure 6.14: Divergence estimated in stabilized (left) and non-stabilized (right) operating conditions.

7 Conclusion

The theoretical and experimental data presented in this thesis are aimed at demonstrating two main points: i) the optimal oculo-motor response for gaze stabilization (i.e. eye velocity, G_{vor}) depends on geometrical parameters of the eye-head system and distance of fixation point; ii) the use of inertial information introduces facilitation in the visual processing.

In subsection 6.2.2 an expression of the amount of residual ocular velocity (measured on the image plane) for different values of G_{vor} gains has been presented. In table 7.1 this expression is used to derive estimates of the average image velocities on the image plane in the following two cases: gaze stabilization is performed using only visual information ($G_{vor} = 0$) and gaze stabilization is performed using inertial information ($G_{vor} = 0.9$). In the first case, the table shows that the range of motion/external disturbances the system can effectively deal with, is limited by the visual measurement process [1]. A saturation effect will cause insufficient oculo-motor compensation and consequently inefficient stabilization even for non-high values of inertial stimulation (see table 7.1 for $\dot{\omega}_h = 20\ deg/s$). The integrated approach, in contrast, seems much more effective and better suited to deal with a wider range of motion or disturbances, overcoming the limitations of the pure visual stabilization.

The data presented in the experimental results are, of course, dependent on the hardware and software components used, and particularly, on the kind of sensors and optical flow estimation algorithm. However, the constraints introduced by these choices are indeed general, and we believe that, the results obtained in this thesis could be generalized to other robotic visual systems. For example, it is conceivable that all algorithms adopted for optical flow estimation do indeed have a limited range of image velocities that can be reliably measured and consequently pure visual stabilization performance is also bounded. Whatever is this range, inertial stabilization improves the performance of the system.

[1] image sampling rate, image processing latency (algorithm complexity, hardware available).

Head velocity (deg/s)	ROV (G_{vor}=0.0) (pixels/frame)	ROV (G_{vor}=0.9) (pixels/frame)
2	0.8	0.1
5	2.1	0.4
10	4.2	0.8
15	6.3	1.3
20	- saturation -	1.7

Table 7.1: Image velocities for fixation point at distance **d** = 100 cm. Frame rate is 12.5 Hz (corresponding to our experimental 80 msec control cycle). Focal length value measured in pixels is approximately 260.

Another important point to stress, is the different latencies of the two "sensory" systems. Latencies are related to the amount of sensor-generated data (much larger in the case of visual sensors), to computational complexity of algorithms used to extract useful sensory cues and to the available hardware. Inertial data are quantitatively some orders of magnitude less than visual data. For example, with 3 rotational and 3 linear accelerometers sampled at 1,000 Hz (which is certainly much more than what the human visual system does), the amount of data that needs to be processed per second is almost negligible with respect to the amount of visual data that need to be processed for reliable optical flow estimation. As a direct consequence, gaze stabilization based on visuo-inertial sensory information, is more responsive: the number of sensorial sources is increased, but to some extent, the computational resources required are optimized, and as a whole, reduced. Furthermore, the inertial data do not depend on visual processing (and vice versa) and their integration does indeed add a completely new and independent data source. This is particularly true in case of external disturbances where even the proprioceptive information measuring head rotation is not relevant. In this case inertial information seems the only source of extra-retinal signals and a powerful tool to increase the overall performance of the system.

The limitations of the present work are, on one side, the use of just one inertial sensor (and consequently the stabilization is limited to one degree of freedom), on the other, the rather simple approach adopted to integrate visual and inertial information and the inappropriate software and hardware architecture, not allowing any multi-tasking, multi-rate visuo-inertial servoing. These issues deserve further investigation. One last point worth mentioning is the extension to a binocular system and the possible use of disparity and proprioceptive eye-angle information to tune the gain of the inertial loop [Biguer and Prablanc, 1981; Snyder and King, 1992; Hine and Thorn, 1987; Sejnowski and Coenen, 1996]. This aspect is currently investigated and may prove even more interesting than simply integrating inertial and optical flow information. As it was shown in section 6.5, the use of inertial information produces a more robust estimation of the expansion component of the optical flow. As a consequence, the vergence control scheme presented in Capurro et. al. [Capurro *et al.*, 97] (which proposes optical flow divergence, computed in the plane of fixation, as the component of retinal velocity driving a fast vergence control loop [Miles, 1995; Busettini *et al.*, 1996; Masson *et al.*, 1997; Busettini *et al.*, 1997]), could be integrated effectively in the system presented in this thesis. The use of the vergence angle (and disparity information) to tune the gain of the VOR loop is also subject of future investigation.

A Appendix

A.1 Log-polar sampling

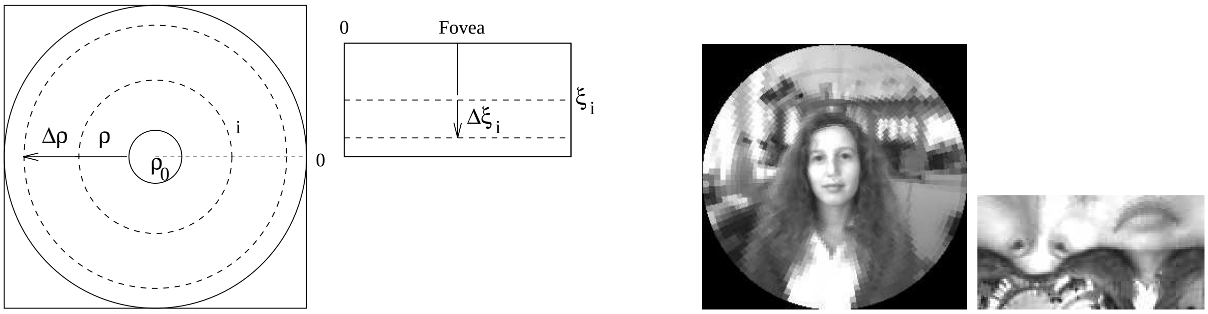
The choice of image representation reveals to be an important issue when a robot must react to environmental stimuli relying only on its visual capability. In many applications a fast reaction, a robust behavior and a low cost implementation are desired characteristics. In order to deal with such requirements, a robot has to refine its visual sensing and starts the process of information selection at the lowest possible level. Among space-variant sensing approaches the log-polar one is receiving a considerable increase of interest [Weiman and Juday, 1990; Sandini and Tistarelli, 1992], [Tunley and Young, 1994; Ferrari *et al.*, 1995b]. The initial analytical formulation of log-polar transformation based on biological data is due mainly to Schwartz [Schwartz, 1977; Schwartz, 1980]. His model described the experimental measures of the mapping between the retina and visual cortex of monkeys [Daniel and Whitteridge, 1961], [Cowey, 1964], [Allman and Kaas, 1971], [Hubel and Wiesel, 1977] as a mapping between a polar plane (ρ, θ) (retinal plane) and a Cartesian plane (η, ξ) (log-polar or cortical plane). The derived logarithmic-polar law is given by the equations (A.1) and (A.2) for the continuous and discrete case respectively:

$$\begin{cases} \eta = q\ \theta \\ \xi = \ln_a \frac{\rho}{\rho_0} \end{cases} \tag{A.1}$$

$$\begin{cases} \eta = q\ \theta_j & j \in [1, .., N_{ang}] \\ \xi = \ln_a \frac{\rho_i}{\rho_0} & i \in [1, .., N_{circ}] \end{cases} \tag{A.2}$$

In the above expressions ρ_0 is the radius of the innermost circle, $1/q$ corresponds to the minimum angular resolution of the log-polar layout (see figure A.1) and (ρ, θ) are the polar coordinates. They are related to the conventional Cartesian reference system by:

$$\begin{cases} x = \rho \cos\theta \\ y = \rho \sin\theta \end{cases} \tag{A.3}$$



(a) (b) (c) (d)

Figure A.1: Parameters of retino-cortical mapping. (a) Any position P in the retinal plane can be expressed in terms of (ρ, θ) or (x, y) coordinates. (b) In the log-polar plane the same position is identified by (η_i, ξ_i). (d) Example of log-polar image. (c) Log-polar image remapped onto Cartesian plane.

At present, the LIRA-Lab binocular vision system performs the log-polar transformation at frame rate using remapping software routines. Hardware remappers [Fisher and Juday, 1988] and prototypes [Debusschere *et al.*, 1989; Rojer and Schwartz, 1990] of space variant CCD's have already been designed and manufactured in the last years and a compact camera using C-MOS technology is already available [Scheffer *et al.*, 1996; Sandini and et.al., 1996]. It is worth noting that the new camera hardware provides a *true* log-polar image: problems of the current software implementation deriving from oversampling in the foveal area and conversely from lack of interpolation in the periphery, are definitively overcome.

A.2 First-order optic flow estimation method

The affine model of the optic flow [Koenderink and van Doorn, 1991] can be described on the basis of four quantities: image *translation*, *rotation*, *divergence* and *shear*. The first two terms specify respectively a rigid two-dimensional translation and rotation of the fixated object. The third term describes an isotropic expansion (or contraction) that specifies a change in scale or a pure deformation. The shear term results in a distortion of the image pattern and corresponds to an expansion in a specified direction and a simultaneous contraction in the perpendicular one in such a way that the area of the pattern is preserved. The component of the affine motion relevant for our measurements is *translation*. The first-order approximation of optic flow (u, v) can be written as:

$$\begin{bmatrix} u \\ v \end{bmatrix} \cong \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} D+S_1 & S_2-R \\ R+S_2 & D-S_1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \tag{A.4}$$

where u_0 and v_0 are the uniform, zero-order components of the flow at the center of the image, D is *dilation*, R, *rotation*, and S_1 and S_2 are the components of *shear*. The equation (A.4) can be re-arranged by taking into consideration the Cartesian to log-polar image velocity transformation and the brightness constancy assumption [Horn, 1986], thus resulting in:

$$-\dot{E} = \begin{bmatrix} \gamma_1 & \gamma_2 & g_\xi & g_\eta & \gamma_3 & \gamma_4 \end{bmatrix} \cdot \begin{bmatrix} u_0 & v_0 & D & R & S_1 & S_2 \end{bmatrix}^T \tag{A.5}$$

where $g_\xi = \frac{\rho_0}{\ln a} \frac{\partial E}{\partial \xi}$, $g_\eta = q\ \rho_0 \frac{\partial E}{\partial \eta}$ and

$$\begin{aligned} \gamma_1 &= \frac{g_\xi \cos\theta - g_\eta \sin\theta}{\rho} , \\ \gamma_2 &= \frac{g_\xi \sin\theta + g_\eta \cos\theta}{\rho} , \\ \gamma_3 &= g_\xi \cos 2\theta - g_\eta \sin 2\theta , \\ \gamma_4 &= g_\xi \sin 2\theta + g_\eta \cos 2\theta , \end{aligned}$$

ρ_0, $1/q$, being the log-polar layout parameters, ρ, θ the polar coordinates.

Explicit knowledge of partial derivatives $\frac{\partial E}{\partial \eta}$, $\frac{\partial E}{\partial \xi}$ and temporal derivative $\frac{\partial E}{\partial t}$ in at least six points of the image is required to solve the system (A.5) using a least square method.

List of tables

List of figures

Bibliography

[A. Buizza *et al.*, 1980] J. Droulez A. Buizza, A. Leger, A. Berthoz, and R. Schmid, "Influence of otolithic stimulation by horizontal linear acceleration on optokinetic nystagmus and visual motion perception," *Exp Brain Res*, 39:165–176, 1980.

[A. Buizza and Droulez, 1981] R. Schmid A. Buizza and J. Droulez, "Influence of linear acceleration in oculomotor control.," In Fuchs and Backer, editors, *Progress in Oculomotor Research*. Elsevier, North Holland, 1981.

[A.A. Skavenski and Winterson, 1979] R.M. Steinman A.A. Skavenski, R.M. Hansen and B.J. Winterson, "Quality of retinal image stabilization during small natural and artificial body rotations in man.," *Vision Research*, 19:675–683, 1979.

[Algrain and Quinn, 1993] M.C. Algrain and J. Quinn, "Accelerometer Based Line-of-Sight Stabilization Approach for Pointing and Tracking Systems," In *Proc. of Second IEEE Conf. on Control Applications*, Vancouver, 1993. IEEE.

[Allman and Kaas, 1971] J.M. Allman and J.H. Kaas, "representation of the visual field in striate and adjoining cortex of the owl monkey (*Aotus trivirgatus*)," *Brain Res.*, 35:89–106, 1971.

[Aloimonos *et al.*, 1988] J. Aloimonos, A. Bandopadhay, and I. Weiss, *Active vision*, Internation journal of Computer Vision 1(4) 1988, 1988.

[Aloimonos, 1993] Yiannis Aloimonos, *Active Vision as a Methodology*, Lawrennce Erlbaum Associates, 1993.

[Angelaki and Hess, 1995] D.E. Angelaki and B.J. Hess, "Inertial representation of angular motion in the vestibular system of rhesus monkeyus.II. Otolith-controlled transformation that depends on an intact cerebellar nodulus.," *J. Neurophysiology*, 73:1729–1751, 1995.

[B. Cohen and Bender, 1964] J.I. Suzuki B. Cohen and M.B. Bender, "Eye movements from semicircular canal nerve stimulation in the cat," *Ann. Otol. Rhinol. Laringol.*, 73:153–169, 1964.

[Baarsma and Collewijn, 1974] E.A. Baarsma and M. Collewijn, "Vestibulo-ocular and optokinetic reactions to rotations and their interaction in the rabbit.," *J. Physiology*, 238:603–625, 1974.

[Bach-y Rita P., 1971] Collins C.C. Bach-y Rita P., editor, *Models of oculomotor neural organization*, Academic Press, 1971.

[Ballard and Brown, 1992] D.H. Ballard and C. M. Brown, "Principles of Animate Vision," *CVGIP*, 56(1):3–21, July 1992.

[Barshan and Durrant-Whyte, 1995] B. Barshan and H.F. Durrant-Whyte, "Inertial sensing for mobile robotics," *Trans. Robotics and Automation*, 11:328–342, 1995.

[Berthoz and Droulez, 1982] A. Berthoz and J. Droulez, "Linear self motion perception.," In A.H. Wertheim, W.A. Wagenaar, and H.W. Leibowitz, editors, *Tutorials on motion perception*. Plenum, 1982.

[Biguer and Prablanc, 1981] B. Biguer and C. Prablanc, "Modulation of the vestibulo-ocular reflex in eye-head orientation as a function of target distance in man.," In A.F. Fuchs and W. Becker, editors, *Progress in oculomotor research*. Elsevier, 1981.

[Blakemore and Donaghy, 1980] C. Blakemore and M. Donaghy, "Co-ordination of head and eyes in the gaze changing behavior of cats.," *J. Physiology*, 300:317–335, 1980.

[Brooks, 1996] R. Brooks, "Behavior-Based Humanoid Robotics," In *Proc. IEEE/RSJ IROS'96*, volume 1, pages 1–8, 1996.

[Buizza and Schmid, 1982] A. Buizza and R. Schmid, "Visual-vestibular interaction in the control of eye movements: mathematical modelling and computer simulation," *Biological Cybernetics*, 43:209–223, 1982.

[Burr *et al.*, 1983] D.C. Burr, J. Holt, J.R. Johnstone, and J. Ross, "Selective depression of motion sensitivity during saccades," *J. Physiology*, 333:1–15, 1983.

[Busettini *et al.*, 1996] C. Busettini, G.S. Masson, and F.A. Miles, "A role for stereoscopic depth cues in the rapid visual stabil ization of the eyes.," *Nature*, 380:342–345, 1996.

[Busettini *et al.*, 1997] C. Busettini, G.S. Masson, and F.A. Miles, "Radial optic flow induces vergence eye movements at ultra-short latencies.," *Nature*, 390:512–515, 1997.

[C. J. Erkelens, 1997] Raymond van Ee C. J. Erkelens, "A computational Model of Depth Perception based on Headcentric Disparity," *Vision Research*, 14:589–592, 1997.

[Capurro *et al.*, 1993] C. Capurro, F. Panerai, and E. Grosso, "The LIRA-lab Head," Technical Report LIRA-Lab TR-3-93, University of Genova, Genova, Italy, September 1993.

[Capurro *et al.*, 97] C. Capurro, F. Panerai, and G. Sandini, "Dynamic Vergence using Log-polar Images," *International Journal of Computer Vision*, 24(1):79–94, 97.

[C.G.Y. Lau *et al.*, 1978] H.A. Jenkins C.G.Y. Lau, V. Honrubia, R.W. Baloh, and R.D. Yee, "Linear Model for Visual-Vestibular Interaction," *Aviation Space Environ. Med*, 49:880–885, 1978.

[Coombs and Brown, 1990] D.J. Coombs and C.M. Brown, "Intelligent Gaze Control in Binocular Vision," In *Proc. of the Fifth IEEE International Symposium on Intelligent Control*, Philadelphia, PA, September 1990.

[Cowey, 1964] A. Cowey, "Projection of the retina on to striate and prestriate cortex in the squirrel monkey, *Saimiri sciureus*," *J. Neurophysiol.*, 27:266–293, 1964.

[Crowley and Bajcsy, 1992] J. Crowley and R. Bajcsy, "ECUS/NFS Workshop on Vision," In *Workshop on Active Vision Hardware*, Ruzenagaard, Sjællands Odde, Denmark, July 1992.

[D. Hyden, 1982] DWF Schwarz D. Hyden, Y. Istl, "Human visual-vestibular interaction as a basis of quantitative clinical diagnostics.," *Acta Otolaryngol.*, 94:53–60, 1982.

[Daniel and Whitteridge, 1961] M. Daniel and D. Whitteridge, "The representation of the visual field on the cerebral cortex in monkeys," *J. Physiol (London)*, 159:203–221, 1961.

[Debusschere *et al.*, 1989] I. Debusschere, E. Bronckaers, C. Claeys, G. Kreider, J. Van der Spiegel, P. Bellutti, G. Soncini, P. Dario, F. Fantini, and G. Sandini, "A 2D Retinal CCD Sensor for Fast 2D Shape Recognition and Tracking," In *Proc. 5th Int. Solid-State Sensor and Transducers*, Montreux, June 25-30 1989.

[Demer and Baloh, 1993] J.L. Demer and R.W. Baloh, "Visual-vestibular interaction in humans during active and passive, vertical head movements," *J. Vestibular Research*, 3:101–114, 1993.

[Demer and Viirre, 1996] J.L. Demer and E.S. Viirre, "Visual-vestibular interaction during standing, walking and running," *J. Vestibular Research*, 6:295–313, 1996.

[Fernandez and Goldberg, 1976] C. Fernandez and J.M. Goldberg, "Physiology of peripheral neurons innervating otolith organs of the squuirrel monkey. I. Response to static tilt and to long-duration centrifugal force.," *J. Neurophysiology*, 39:970–984, 1976.

[Ferrari *et al.*, 1995a] F. Ferrari, J. Nielsen, P. Questa, and G. Sandini, "Space variant imaging," *Sensor Review*, 15(2):17–20, 1995.

[Ferrari *et al.*, 1995b] F. Ferrari, J. Nielsen, P. Questa, and G. Sandini, "Space Variant Sensing for Personal Communication and Remote Monitoring," In *Proc. of EU-HCM Smart Workshop Lisbon (Portugal)*, April 1995.

[Fisher and Juday, 1988] T.E. Fisher and R.D. Juday, "A Programmable Video Image Remapper," In *Proc. SPIE Conf. on Pattern Recognition and Signal Processing*, volume 938, pages 122–128, Orlando, FL, 1988. SPIE.

[Fuchs, 1967] A. F. Fuchs, "Periodic eye tracking in the monkey.," *J. Physiology*, 193:161–171, 1967.

[G.E. Grossman *et al.*, 1988] L.A. Abel G.E. Grossman, R.J. Leigh, D.J. Lanska, and S.E. Thurston, "Frequency and velocity of rotational head perturbations during locomotion," *Exp Brain Res*, 70:470–476, 1988.

[Goldberg and Fernandez, 1975] J.M. Goldberg and C. Fernandez, "Responses of peripheral vestibular neurons to angular and lienar acceleration in the squirrel monkey," *Acta Otolaryngol.*, 80:101–110, 1975.

[G.R. Barnes and Eason, 1987] S.F. Donnelly G.R. Barnes and R.D. Eason, "Predictive velocity estimation in the pursuit reflex response to pseudo-random and step displacement stimuli in man.," *J. Physiology*, 389:111–136, 1987.

[Gregory *et al.*, 1995] Gregory, Harris, Heard, and Rose, *The Artful Eye*, Oxford University Press, 1995.

[H. Collewijn and Takahashi, 1972] C.W. Oyster H. Collewijn and E. Takahashi, "Rabbit optokinetic reactions and retinal direction-selective cells. A preliminary model," *Bib. Ophtalm.*, 82:280–287, 1972.

[Hine and Thorn, 1987] T. Hine and F. Thorn, "Compensatory ete movements during active head rotation for near targets: effects of immagination, rapid head oscillation and vergence," *Vision Research*, 27:1639–1657, 1987.

[Horn, 1986] B. K. P. Horn, *Robot Vision*, MIT Press, Cambridge, USA, 1986.

[Hubel and Wiesel, 1977] D.H. Hubel and T.N. Wiesel, "Functional architecture of macaque monkey cortex," *Proc. R. Soc. Lond.*, 198:1–59, 1977.

[K. E. Money, 1968] M. J. Correia K. E. Money, "The Vestibular System of the Owl," In *Proc. of the Fourth Symposium on the Role of Vestibular Organs in Space Exploration*, Pensacola, FL - USA, September 1968.

[Kandel *et al.*, 1991] E.R. Kandel, J.H. Schwartz, and T.M. Jessel, *Principles of Neuroscience*, Elsevier, 1991.

[Keller, 1974] E.L. Keller, "Participation of the medial pontine reticular formation in eye movement generation in monkey.," *J. Neurophysiology*, 37:316–332, 1974.

[Keller, 1978] E.L. Keller, "Gain of vestibulo-ocular reflex in the monkey at high rotational frequencies.," *Vision Research*, 18:311–315, 1978.

[Koenderink and van Doorn, 1991] J. Koenderink and J. van Doorn, "Affine Structure from Motion," *Journal of the Optical Society of America*, 8(2):377–385, 1991.

[Lobo and Dias, 1997] J. Lobo and J. Dias, "Towards visual and inertial sensing integration: contribution to improve mobile robot autonomy," In *Proc. 5th Symposium on Intelligent Robotics Systems*, Stockholm, Sweden, July 1997.

[Malcolm and Melvill-Jones, 1970] R. Malcolm and G. Melvill-Jones, "A quantitative stady of vestibular adaptation in humans," *Acta Otolaryngol.*, 70:126–135, 1970.

[Masson *et al.*, 1997] G.S. Masson, C. Busettini, and F.A. Miles, "Vergence eye movements in response to binocular disparity wit hout the perception of depth.," *Nature*, 389:283–286, 1997.

[Mayne, 1965] R. Mayne, "The Match of the Semicircular Canals to the Dynamic Requirements of Various Species," In *Proc. of the First Symposium on the Role of Vestibular Organs in Space Exploration*, NASA SP-77, - 1965.

[Melivil-Jones and Spells, 1963] G. Melivil-Jones and K.E. Spells, "A theoretical and comparative study of the functional dependence of the semicircular canals upon its physical dimensions.," *Proc. R. Soc. B.*, 157:403–419, 1963.

[Melvill-Jones, 1969] G. Melvill-Jones, "The function significance of semicircular canal size.," In H.H. Kornhuber, editor, *Handbook of Sensory Physiology*. Springer Verlag, 1969.

[Melvill-Jones, 1971] G. Melvill-Jones, "Organization of neural control in the vestibulo-ocular reflex arc.," In Collins C.C. Bach-y Rita P., editor, *The Control of eye movements*. Academic Press, 1971.

[Melvill-Jones and Milsum, 1965] G. Melvill-Jones and J. H. Milsum, "Neural Dynamics of Vestibular Trsnsduction," In *Proc. of the 18th Ann. Conf. on Eng. in Med. and Biol.*, volume 7, pages 38–39, USA, 1965.

[Micheal and Melvill-Jones, 1966] J.A. Micheal and G. Melvill-Jones, "Dependence of visual tracking capability upon stimulus predictability.," *Vision Research*, 16:707–716, 1966.

[Miles, 1995] F. A. Miles, "The sensing of optic flow by the primate optokinetic system," In R.W. Kentridge J.M. Findlay and R. Walker, editors, *Eye Movement Research: Mechanism, Processes and Applications.* Elsevier Press, 1995.

[Morasso *et al.*, 1973] P. Morasso, E. Bizzi, and J. Dichgans, "Adjustment of saccade characteristics during head movements," *Exp. Brain Res.*, 16:492–500, 1973.

[Murphy, 1978] B.J. Murphy, "Pattern thresholds for moving and stationary gratings during smooth eye movement.," *Vision Research*, 18:521–530, 1978.

[Nakayama, 1983] K. Nakayama, "Motion parallax sensitivity and space perception.," In A.Hein and M. Jeannerod, editors, *Spatially oriented behavior.* Springer-Verlag, 1983.

[Ogle, 1964] K.N. Ogle, *Researches in Binocular Vision*, Hafner Publishing Company, London, 1964.

[Pahlavan *et al.*, 1992] K. Pahlavan, Tomas Uhlin, and Jan-Olof Eklundh, "Integrating Primary Ocular Processes," In *Proc. ECCV92 - European Conference of Computer Vision*, volume LNCS-588, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.

[Paige, 1991] G. D. Paige, "Linear Vestibulo-Ocular Reflex (LVOR) and Modulation by Vergence," *Acta Otolaryngol Suppl*, 481:282–286, 1991.

[Panerai and Sandini, 1997] F. Panerai and G. Sandini, "Integration of Inertial and Visual Information in an Active Vision System," Technical Report 4/97, LIRA-Lab, University of Genoa, Genova, Italy, April 1997.

[P.D. Poulaski, 1981] D.A. Robinson P.D. Poulaski, D.S. Zee, "The behavior of the vestibulo-ocular reflex at high velocities of head rotation.," *Brain Res*, 222:159–165, 1981.

[Questa and Sandini, 1996] P. Questa and G. Sandini, "Time to contact computation with a space-variant retina-like C-mos sensor," In *Proc. Int. Conference on Intelligent Robots and Systems*, Osaka - Japan, November 1996. JRS-IEEE.

[R.D. Tomlinson, 1980] D.W. Schwarz R.D. Tomlinson, G.E. Saunders, "Analysis of human vestibulo-ocular reflex during active head movements.," *Acta Otolaryngol.*, 90:184–190, 1980.

[Robinson, 1977] D. A. Robinson, "Vestibular and optokinetic symbiosis: an example of explaining by modelling.," In R. Baker and A. Berthoz, editors, *Control of gaze by brain stem neurons.* Elsevier/North Holland Biomedical Press, 1977.

[Robinson and Peterson, 1985] D.L. Robinson and S.E. Peterson, "Response of pulvinar neurons to real and self-induced stimulus motion," *Brain Res*, 338:392–394, 1985.

[Rogers and Graham, 1982] B. Rogers and M. Graham, "Similarities between motion parallax and stereopsis in human depth perception," *Vision Research*, 22:261–270, 1982.

[Rojer and Schwartz, 1990] Alan S. Rojer and Eric L. Schwartz, "Design Considerations for a Space-Variant Visual Sensor with Complex-Logarithmic Geometry," In *Proc. 10th ICPR*, pages 278–285. IEEE Computer Society, Atlantic City, June 1990.

[Sandini and et.al., 1996] G. Sandini and et.al., "Image-based Personal Communication Using an Innovative Space-variant CMOS Sensor," In *Proc. Roman '96*, Tsukuba, Japan, November 1996.

[Sandini and Tistarelli, 1992] G. Sandini and M. Tistarelli, *Neural Network for Perception*, chapter Vision and space variant sensing, pages 398–425, Academic Press, 1992.

[Sargent and Paige, 1991] E.W. Sargent and G.D. Paige, "The primate vestibulo-ocular reflex during combined linear and angular head motion," *Exp Brain Res*, 87:75–84, 1991.

[Scheffer *et al.*, 1996] D. Scheffer, B. Dierickx, and F. Pardo, "Log-polar Image Sensor in CMOS Technology," In *Proc. Europto*, Besancon, June 10-14 1996.

[Schwartz, 1977] E. L. Schwartz, "Spatial Mapping in the Primate Sensory Projection: Analytic Structure and Relevance to Perception," *Biological Cybernetics*, 25:181–194, 1977.

[Schwartz, 1980] E. L. Schwartz, "A Quantitative Model of the Functional Architecture of Human Striate Cortex with Application to Visual Illusion and Cortical Texture Analysis," *Biological Cybernetics*, 37:63–76, 1980.

[Schwarz and Miles, 1991] U. Schwarz and F. Miles, "Ocular Responses to Translation and their Dependence on Viewing Distance. I. Motion of the Observer," *J. Neurophysiology*, 66:851–864, 1991.

[Sejnowski and Coenen, 1996] T.J. Sejnowski and O. J.M.D. Coenen, "A dynamical model of context dependancies for the vestibulo-ocular reflex," In M. Mezer D. Touretzky and M. Hasselmo, editors, *Advances in Neural Information Processing Systems 8*. MIT Press, 1996.

[Sharkey, 1995] P.M. Sharkey, "Bandwidth issues for active vision systems," In *Int. Conf. Mechatronics and Machine Vision in Practice*, Hong Kong, Sept 1995.

[Sharkey *et al.*, 1995] P.M. Sharkey, D.W. Murray, and P.F. MacLauchlan, "A mechatronic approach to active vision systems," In *Int. Conf. Mechatronics and Machine Vision in Practice*, Hong Kong, Sept 1995.

[Shioiri and Cavanagh, 1989] S. Shioiri and P. Cavanagh, "Saccadic suppression of low-level motion," *Vision Research*, 29:915–928, 1989.

[Snodderly and Kurtz, 1985] D.M. Snodderly and D. Kurtz, "Eye position during fixation tasks:comparison of macaque and human.," *Vision Research*, 25:83–98, 1985.

[Snyder and King, 1992] L.H. Snyder and W.M. King, "Effects of viewing distance and location of the axis of head rotation on the monkey's vestibuloocular reflex I. Eye movement response," *J. Neurophysiology*, 67:861–874, 1992.

[Steinman and Levinson, 1990] R.M. Steinman and J.Z. Levinson, "The role of eye movements in the detection of contrast and spatial detail," In E. Knowler, editor, *Eye movements and their role in visual and cognitive processes*. Elsevier, 1990.

[Steinman *et al.*, 1985] R.M. Steinman, J.Z. Levinson, H. Collewijn, and J. Van der Steen, "Vision in the presence of known natural retinal image motion," *J. Opt Soc. Am.*, 2:226–233, 1985.

[Sundareswaran, 1991] V. Sundareswaran, "Egomotion from Global Flow Field Data," In *Proc. of the IEEE Workshop on Visual Motion*, Princeton, NJ - USA, October 1991.

[T. Raphan and Matsuo, 1977] B. Cohen T. Raphan and A. Matsuo, "A velocity storage mechanism responsible for optokinetic nystagmus (OKN), optokinetic after-nystagmus (OKAN) and vestibular nystagmus," In R. Baker and A. Berthoz, editors, *Control of gaze by brain stem neurons*. Elsevier/North Holland Biomedical Press, 1977.

[Tunley and Young, 1994] H. Tunley and D. Young, "First order optical flow from log-polar sampled images," In *Proc. Third European Conference on Computer Vision*, May 2-6 1994.

[U. Schwarz and Miles, 1989] C. Busettini U. Schwarz and F. A. Miles, "Ocular Responses to Linear Motion are Inversely Proportional to Viewing Distance," *Science*, 245:1394–1396, 1989.

[Vieville and Faugeras, 1990] T. Vieville and O. D. Faugeras, "Computation of Inertial Information on a Robot," In *Proc. of Fifth International Symposium of Robotic Research*, Tokio, 1990. MIT Press.

[von Seelen and Madden, 1992] U. Cahan von Seelen and B. Madden, “Binocular Camera Platform Home Page,” Technical Report http://www.cis.upenn.edu/ grasp/head/headpage/headpage.html, GRASP Lab - Univ. of Pensylvania, Philadelpia, USA, 1992.

[Waespe and Henn, 1977] W. Waespe and V. Henn, “Neuronal activity in the vestibular nuclei of the alerto monkey during vestibular and optokinetic stimulation,” *Exp Brain Res*, 27:523–538, 1977.

[Weiman and Juday, 1990] C. F. R. Weiman and R. D. Juday, “Tracking Algorithms Using Log-Polar Mapped Image Coordinates,” In *SPIE Int. Conf. on Intelligent Robots and Computer Vision VIII: Algorithms and Techniques*, volume 1192, pages 843–853, Philadelphia (PA), 6-10 November 1990.

[Weiman, 1995] C.F.R. Weiman, “Binocular Stereo via log-polar retinas,” In *Proc. SPIE AeroSense95*, Orlando, Florida, April 1995.

[Westheimer and McKee, 1975] G. Westheimer and S.P. McKee, “Visual acuity in presence of retinal image motion,” *J. opt. Soc. Am.*, 65:847–850, 1975.

[Wilson and Melvill-Jones, 1979] V.J. Wilson and G. Melvill-Jones, *Mammalian vestibular physiology*, Plenum Press, 1979.

[Young and Meiry, 1966] L. R. Young and J. L. Meiry, “A Revised Dynamic Otolith Model,” In *Proc. of the Second Symposium on the Role of Vestibular Organs in Space Exploration*, Moffet Field, CA - USA, January 1966.

[Young and Oman, 1969] L. R. Young and C. M. Oman, “Model for vestibular adaptation to horizontal rotation,” *Aerosp. Med.*, 40:1076–1080, 1969.

[Young and Oman, 1968] L.R. Young and C.M. Oman, “A model for vestibular adaptation to horizontal rotation,” In *Proc. of the Fourth Symposium on the Role of Vestibular Organs in Space Exploration*, Pensacola, FL - USA, September 1968.