

# Biologically inspired Gaze Stabilization

*A first step toward live visual Object–Recognition  
in Walking Robots*

H.J. van de Zedde, 1014978

*December 2002*

*A thesis submitted for the degree of master of  
Artificial Intelligence at the University of Groningen*

*Supervisors:*

Prof.Dr. Lambert Schomaker	University of Groningen
Prof. Dr. Rolf Pfeifer	University of Zürich
Fumiya Iida	University of Zürich

# Abstract

*Gaze stabilization* is the process in which the image projected on the retina is kept stationary. The goal of this thesis is to give an outline of how other species solve this problem and what mechanisms are involved in this behavior. Based on these findings a biologically plausible model, the so-called *Elementary Motion Detector* or *Reichardt Detector* (Reichardt, 1969; Borst and Egelhaaf, 1993; Iida, 2001), is proposed and implemented in a walking robot dog, which has a novel musculo-skeletal design based on anatomical studies of the canine.

This research project focuses on the analysis of the optokinetic reflex. The model is a method to measure the amount of retinal slip, the displacement of the surrounding environment of the world. This model is implemented in a closed-loop manner in an artificial robot dog. The output of the model generates compensatory eye movement signals to control the eyes of the dog. The performance of this model is tested in a real-world office environment.



# Acknowledgements

## Zürich, Switzerland

The final stage of my study of Artificial Intelligence was a graduation project, consisting of a six month research period. I preferred a project abroad, and arranged an undergraduate position at the AI-lab, university of Zürich. Hereby I want to show my gratitude to Prof. Dr. Rolf Pfeifer, the leading professor at the lab, for giving me the opportunity to work in his lab.

The robot dog project was initiated by Fumiya Iida, a PhD-student at the AI-lab in Zürich and my supervisor. I would like to thank Fumiya for his enormous time-investments in realizing the hardware aspects of the robot dog. So that I was able to do my experiments and achieve the results presented in this thesis. Thanks also to Gert Kootstra for giving me many good ideas and definitely for the great time in Zürich. Furthermore I would like to thank my local supervisor prof. dr. Lambert Schomaker for his useful suggestions in the final stage of my thesis, Sarah and Bartje for reading and correcting my thesis and finally thanks to all the AI-lab members.



# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Research Questions . . . . .	2
1.2	The AI-lab in Zürich, Switzerland . . . . .	3
1.3	Thesis Outline . . . . .	4
<b>2</b>	<b>How is a Stable Image realized in Biological Systems?</b>	<b>5</b>
2.1	Stabilizing Reflexes . . . . .	5
2.2	Gaze Stabilization in Biological Systems . . . . .	8
2.3	Three-dimensional Projection of the Vestibular and Visual Signals . . . . .	14
2.4	Characteristics of Optical Flow . . . . .	15
2.5	What does Biological Information teach us . . . . .	16
2.6	Conclusion . . . . .	18
<b>3</b>	<b>Gaze stabilization in existing Artificial Systems</b>	<b>19</b>
3.1	Shibata and Schaal . . . . .	19
3.2	The Babybot Project . . . . .	20
3.3	Cog Project . . . . .	20
3.4	The Kismet Project . . . . .	21
<b>4</b>	<b>A Biologically Plausible Model to Detect Image Motion</b>	<b>23</b>
4.1	Processing of Optic Flow in Biological Systems . . . . .	23
4.2	Development of the Elementary Motion Detector . . . . .	24
4.3	Implementation of the Motion Detector . . . . .	29
4.4	The Extended Elementary Motion Detector . . . . .	32
4.5	Discussion . . . . .	34
<b>5</b>	<b>Gaze Stabilization in the Robot Dog</b>	<b>35</b>
5.1	The Robot Dog . . . . .	35
5.2	Experimental Results . . . . .	37
5.3	Gaze Stabilization Results . . . . .	44
5.4	Discussion . . . . .	45
<b>6</b>	<b>Conclusion</b>	<b>47</b>
<b>A</b>	<b>Pictures of the Robot Dog</b>	<b>55</b>
<b>B</b>	<b>Detection Range of the EEMD</b>	<b>57</b>
<b>C</b>	<b>Additional Gaze Stabilization Results</b>	<b>59</b>



# Chapter 1

## Introduction

Every organism has adapted itself to its environment, in a way that increases the chances of surviving and reproducing. Surviving is a matter of coping with sudden changes in the environment; a predator can appear suddenly or a beautiful woman can pass by. But in order for its movements to be regulated by the environment, it must be able to detect structures and events in its surroundings. Several mechanisms have been developed by evolution to perceive information about the environment. For instance abilities to detect smell, pressures on the body surface, forces on the limbs and muscles, sound waves in the air (or in water) and in some instances, to sense electric and magnetic fields are a few examples.

An animal, which is able to smell, can detect the presence of nearby food or predators, but is it also able to pinpoint their location based on smell only? The ability to perceive pressures on the body surface and forces on the limbs and muscles only give information about the animals direct surrounding, and do not supply anything about the environment three meters away for instance. Sound can provide information about distant animals, but mostly, except for example in bats, sound tells nothing about inanimate structures present in the environment.

Thus sensitivity to these signals gives an animal considerable perceptual abilities, but leaves it unable to detect information rapidly about either its animate surroundings or about a silent animal, prey or predator at a distance from itself. Therefore animals have developed the sensitivity for light, the ability to perceive their surroundings through vision. Most animals use this information source thoroughly, although some species are known to move about rapidly without vision, such as bats, dolphins and other cetaceans (Bruce, Green, and Georgeson, 1996). These species use an echolocation system based on ultrasonic cries. Some fish species living in murky water, detect objects in their surroundings by distortion of their own electric fields.

The overall conclusion is that visual perception of the surrounding world is an important source of information for the survivability of an animal. Therefore it is extremely relevant to use this source as efficiently as possible. Through vision an animal is able to detect the 3D-structure of its surrounding, food from a distance, a predator or prey, where to put its feet while walking etc. But a problem arises because the visual sensors, the eyes, are attached to the head. The problem of gaining information from this sensor is that motion, either elicited by the surrounding or by the animal itself, for instance ego-motion caused by running, disturbs the signals. If the head moves the eyes are also moving. Imagine that you are jumping. As you will notice, your

eyes are making stabilizing movements. If this were not the case you would not be able to perceive anything sensible from the wildly moving projections of the surroundings on your retina. Therefore the basic task of this visual system basic task, before successfully-perceiving the world, consists of dealing with the problem of maintaining a stable visual representation during ego-motion.

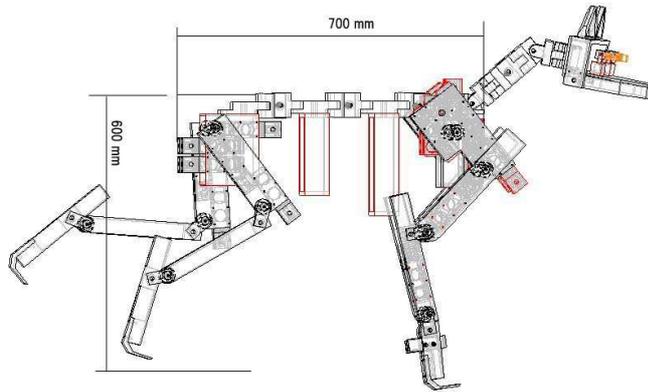


Figure 1.1: *The mechanical robot dog design*

## 1.1 Research Questions

This graduation project had the goal to give an answer to the following three questions:

- **What are the mechanisms behind gaze stabilization during ego-movements?**
- **How can these mechanisms be realized in artificial systems?**
- **How does the proposed gaze stabilization help higher cognitive tasks such as object recognition?**

As we will see in this thesis, not only visual signals are used in *gaze stabilization*. Nature came up with several mechanisms to cope with the disturbance of the visual signal by ego-motion. Every species has its own specialized solution.

In building walking and running robots, *gaze stabilization* is a feature that must be implemented, because otherwise the robot suffers from translational and rotational image flow. Biological organisms have been dealing with this problem for millions of years, and evolution has fine tuned their entire body structure and functionalities to optimize their survival chances. Inspired by the efficient locomotion of a biological organism like the dog, a project was initiated by Fumiya Iida, a PhD-student at the AI-lab. The objective of this project was to understand the perceptive processes which are the consequence of the physical structure of the dog. In the design of the dog (Fig. 1.1), we tried to imitate the physical properties; the proportion, the skeleton's weight, number and positions of joints and the location of the actuators (muscles).

The first goal of the project is to realize locomotion. Locomotion is a crucial feature for an animal, the project's aim is to build a robot that mimics the efficient locomotion of a canine, for example in walking, running and jumping (see fig. 1.2 <sup>1</sup>). The second goal of the project is to understand sensory-motor control. The first focus was the coupling of visual input to the eye motors to provide a stable view of the world while the dog is jumping, running or walking. Therefore the gaze of the robot dog needed stabilization. In this project we have analyzed in what way *gaze stabilization* is obtained in the original biological organism, the dog itself. We created a model of how motion is perceived by the eyes of a biological organism and how we could use this signal to send stabilizing signals to the eyes of the robot dog.

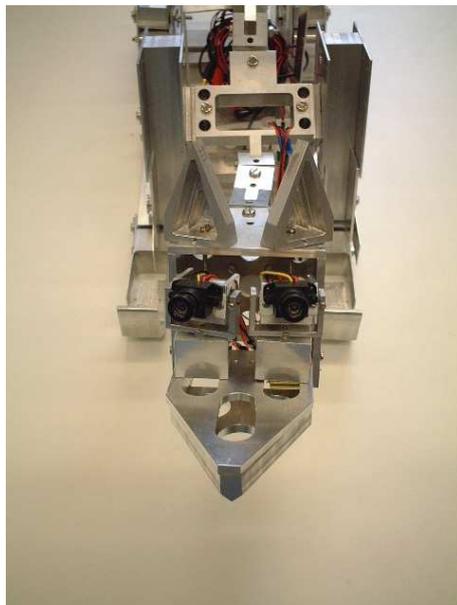


Figure 1.2: *The Real Robot Dog*

## 1.2 The AI-lab in Zürich, Switzerland

At the AI-lab research is conducted with a very clear philosophy. The researchers at the AI-lab aim, not only to create systems that produce intelligent behavior, but to comprehend the principles that underlie this behavior in other robust systems. Good examples of robust autonomous systems are found in nature (humans, animals and insects). Research at the lab is focussed on these systems. A good understanding is reached by analyzing these natural forms of intelligence and understanding their efficient way of problem solving. Inspired by these findings, artificial intelligent systems are designed and built. This results first of all in proving the assumed model in the biological system and second in the creation of a robust artificial system.

Instead of the analytic approach, universally applied in all empirical sciences, the AI-lab uses a synthetic methodology. This methodology operates by creating an artificial system that reproduces certain aspects of a natural system. Instead of focusing

---

<sup>1</sup>Appendix A contains more robot dog pictures

on merely producing the correct experimental results, i.e., obtaining the correct output, this methodology strives to understand the internal mechanism that produced these particular results. The discipline that uses this methodology is called “*embodied cognitive science*” (Pfeifer and Scheier, 1999). This approach can be characterized as “*Understanding by building*”. The interaction between biology and autonomous agents research is interesting. For example, we may want to replicate an idea from nature, say, the path finding behavior of an ant. After defining the internal model of this path-finding mechanism, the proof that the model could be correct can be obtained by creating an artificial agent, on which this model is implemented. By analyzing the behavior of this artificial agent and comparing it to the natural agent, conclusions can be drawn about the correctness of the proposed model. In the AI-lab, biologists and psychologists learn from building robots and developing computer programs, and engineers and computer scientists can learn about nature. The AI-lab consists, then, of a large diversity of backgrounds. This is based on the conviction that the interaction between various disciplines is highly productive.

### 1.3 Thesis Outline

The thesis is structured as follows; in chapter two we give an overview of the mechanisms behind gaze stabilization in biological systems. The optokinetic reflex (OKR), the reflex that stabilize the eyes based on visual information, the vestibulo-ocular reflex (VOR), the reflex that relies on signal from the vestibular organ and some other assisting reflexes are discussed in this chapter. Next, an overview is presented of several species and their gaze stabilization techniques. Then we present the characteristics of optical flow. In the final section of chapter two we discuss the potential fruitful cooperation between biologists and engineers. In chapter three gaze stabilization in existing artificial systems will be treated. After that a biologically inspired motion detection model based on the results from biological research performed by Reichardt in insects will be developed in chapter four. The pros and cons of this model and its implementation will be discussed. In chapter five we present the gaze stabilization results obtained from the robot dog. And with the conclusion we finish this thesis.

## Chapter 2

# How is a Stable Image realized in Biological Systems?

*In the first section we will analyze the reflexes that cooperate to maximize the performance of gaze stabilization. First the two dominant reflexes are treated: the optokinetic reflex (OKR) and vestibulo-ocular reflex (VOR), followed by a section that gives an overview of the other supporting reflexes. We then present in what way biological systems gain gaze stabilization. We mention insects, birds, chameleons, rabbits and finally primates. After that, we analyze the characteristics of optical flow and we close this chapter with a discussion about what this biological information teach us in building artificial intelligent systems.*

## 2.1 Stabilizing Reflexes

*Gaze stabilization* is a mechanism, driven by several reflexes. According to the English dictionary the definition of a reflex is: “a non-conscious reaction to a nerve stimulation”. Therefore gaze stabilization is a mechanism the animal is not aware of. In other words it cannot control or suppress gaze stabilization, the mechanism is executed automatically.

### 2.1.1 Optokinetic Reflex

As an introduction to the optokinetic reflex, the following personal anecdote:

One day, I was traveling by train from Groningen to Arnhem. I saw a girl in front of me staring out of the window acting really strange with her eyes. Her eyes were focussed on a location outside of the train. While tracking the object her eyes were moving very fast to the left. The moment her eyes almost disappeared in the corner of her eye, the eye jumped back to the middle, and started tracking a new object. I wondered what kind of mechanism was the source of this behavior.

The phenomenon presented here is called the optokinetic reflex. Bárány mentioned it already in 1921 and came up with the word ‘train-nystagmus’<sup>1</sup> (Oey, 1979). In animals with move-able eyes, the oculomotor system generates compensatory eye movements functioning to stabilize the retinal image. Animals whose eyes are fixed, insects for

---

<sup>1</sup>Nystagmus: a rapid, involuntary, oscillatory motion of the eyeball

instance, generally attempt to stabilize the head, as some animals with small heads and flexible necks also do. In all cases the intent is to stabilize the gaze, i.e. the position of the eyes with respect to the surrounding environment.

As an example to show the importance of the gaze stabilizing system, hold your finger in front of your face, and swing your finger slowly from left to right, while looking at this finger. Try to extract the details of an object in the background. As you see, the background is extremely blurred (Howard, 1993). This blurring is an extreme form of retinal slip. As you might have noticed, relying on such an unstable information source is very inconvenient. Without the gaze stabilization mechanism, images on the retina would always be blurred like this. If, for instance, a dog is running or jumping up and down, the world that he visually perceives also bounces on its retina, but because every natural system has its way to cope with this problem, to the dog it will not look like the entire world is bouncing. The dog will make stabilizing movements with the eye/neck muscles, in order to keep the world stationary on the retina. By means of this *gaze stability* system the possibility is created to perceive visual information about events that happen in front of the dog undisturbed.

The optokinetic reflex functions to minimize the retinal slip. Retinal slip is defined as the overall velocity with which an image drifts on the retina (Shibata and Schaal, 2001). The optokinetic reflex relies on visual information coming from the entire retina, and not just the fovea (Draper, 1998). The goal of the OKR is to keep the image still on the fovea, the center of the retina. This process is called 'smooth pursuit'. On the other hand the eye has to make saccades, quick correcting movements to prevent the eye from moving in its corner. This is called '*nystagmus*'. The optokinetic reflex controlling the eyes functions as a visually driven closed-loop negative feedback tracking system, i.e. an output is produced, the eye movement, that operates to reduce the input, the retinal image motion.

### 2.1.2 Vestibulo-ocular Reflex

A short anecdote about the vestibular mechanism:

- Do not move your head, wave your hand in front of your eyes at a moderate frequency, while you try to keep track of the hand. It seems blurred, your eyes are not able to track your hand properly.
- Do not move your hand, turn your head from left to right at a similar frequency, keep an eye on your hand. There is no blurring.  
(Shibata and Schaal, 2001; Burdess, 1996)

Tracking of the hand based on visual information, the first condition of the anecdote, is clearly slower than the second example. In this case information from the vestibular organ, measurement of the head movements, is used to correct the eyes directly. This is an example of the vestibulo-ocular reflex (VOR). The vestibulo-ocular reflex controls the eyes as an open-loop, i.e. output is produced, the eye movement, that does not affect the input, the head movement. The VOR is very rapid and functions best at high frequencies and movements of the head.

The vestibulo-ocular reflex is based on signals produced by the vestibular labyrinth in the middle ear. The task of this vestibular organ is to determine the absolute movements of the head in three dimensional space, i.e. three linear- and three rotational movements. There is one vestibular system on each side of the head. Each system

consists of two types of sensors; the otholith organs, which sense the linear movement; and a set of three semicircular canals arranged at right angles to each other, sensing rotation movement in three planes (Burdess, 1996). The otholith organs are able to sense the orientation and the magnitude of the gravitational vector relative to the head (Tan, 1992). The otholith organs consist of the utricle, lying in the horizontal plane, and the saccule, lying the vertical plane. These two sensors respond to respectively linear horizontal- and vertical forces. By combining these two signals, the third dimension can be extracted.

The three semicircular canals are more or less orthogonal with respect to each other. Rotational angular acceleration of the whole canal causes fluid to be left behind on account of its inertia, which is translated in a signal according to this acceleration. The canal system measures the acceleration, but performs an integration on this signal to extract the velocity signal. Thus the semicircular canal system acts as an angular speedometer. Its neural output is directly proportional to the angular velocity of head movements. By combining the output of the three canals, the brain can create a representation of the vector, which describes the speed of head rotation in three dimensions.

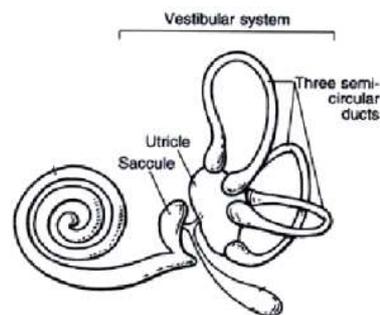


Figure 2.1: *The vestibular System*

### 2.1.3 Supporting Reflexes

Besides these main two reflexes, other reflexes also play a minor role in the gaze stabilization process, although most papers do not mention them (Shibata and Schaal, 2001; Metta, Panerai, and Sandini, 2000; Scassellati, 2001). Therefore in this thesis these reflexes will be mentioned if necessary, otherwise the assumption is that *gaze stabilization* consists only of the optokinetic reflex and vestibular reflex. The three supporting reflexes are briefly described below.

#### Vestibulo-collicular Reflex

This reflex relies on signals from the vestibular organ, but instead of executing stabilizing eye movements, neck movements stabilize the image on the retina. This reflex is clearly used for example in birds with long, flexible necks, as we will see in section 2.2.2 on page 11.

#### Opto-collicular Reflex

Instead of relying on the vestibular organ, this reflex uses the amount of retinal slip to minimize this optic flow with stabilizing neck movements. An example of the opto-collicular reflex is mentioned in section 2.2.2.

#### Cervico-ocular Reflex

The cervico-ocular reflex reacts on signals from the neck muscles with stabilizing eye movements (section 2.2.3). In biological experiments this reflex is elicited by straining the head, while the rest of the body performs sinoidal movements. The function of the *cervico-ocular reflex* is unclear, especially because the eye movements often alternate between compensatory and anti-compensatory movements (Gioanni, Bennis, and

Sansonetti, 1993). This might be the reason that in most *gaze stabilization* literature the cervico-ocular reflex is not mentioned at all.

## 2.2 Gaze Stabilization in Biological Systems

In every species *gaze stabilization* is achieved by a cooperation of multiple reflexes to perceive the world with a minimized retinal slip. These reflexes are mainly the optokinetic reflex (OKR) and the vestibulo-ocular reflex (VOR) assisted by several other reflexes. But in every organism these reflexes are developed in a different manner. In the following sections we present an overview of several animals and their way of obtaining and using the signals to gain gaze stabilization, starting with insects, followed by birds, chameleons, rabbits and finally primates.

### 2.2.1 Gaze Stabilization in Insects

Let's concentrate first on the several sensor types insects have developed to perceive the world.

#### Propioceptors (self-perception receptors)

**Halteres:** The halteres are the evolutionary remnants of a second set of wings. More ancestral insects, like the dragon-fly still possess 2 pair of wings. The haltere is the mechanical equivalent of a gyroscope and this structure is responsible for why flies fly so well. It measures the amount of self-motion during flight (Hengstenberg, 1993).

The halteres relay signals to the wing muscles to alter their stroke or angle. Halteres (fig. 2.2) are small knobs that beat in time, but out of phase of the wings. When the thorax alters direction, the haltere is twisted, and the fly can respond accordingly.

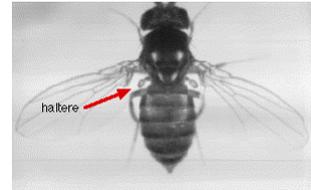
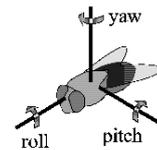


Figure 2.2: *The haltere*

**Hair Plates:** Hair plates are tight clusters of short hairs on sockets. They are found at the articulation of joints or body parts where they are stimulated as the segments move in relation to one another, providing feedback on limb position (Gullan and Cranston, 2000). This is used for instance to measure the gravitational vector on a tilted surface, while the insect is walking (in this state the halteres do not function). The head is move-able to the thorax, in the insect's body by ca. 20 muscles per side. For example, the insect can yaw and pitch its head  $\pm 20$  deg and roll its head  $\pm 90$  degrees (Schilstra, 1999).



#### The Visual System

There are two main considerations in analyzing the visual system of insects: (i) their resolving power for images, i.e. the fine amount of detail that can be resolved; and (ii) their light sensitivity, i.e. the minimum ambient light level at which insects can see (Gullan and Cranston, 2000).

**Dermal detection:** Light detection through their body surface is defined as dermal detection. The insect has sensory receptors below the body cuticle, i.e. the upper skin. An insect with dermal detection would be able to use the light changes detected by the body surface to orient defensive responsive methods according to the rough direction

of approach of a predator (Bruce, Green, and Georgeson, 1996).

**Ocelli:** Typically three small ocelli lie in a triangle on top of the head. Ocelli are small, and have a single, under-focussed wide-angle lens and a few hundred photoreceptors (Hengstenberg, 1993). They are sensitive to low light intensities and to subtle changes in light, but not suited for high-resolution vision. They appear to function as horizon detectors for control of roll and pitch movements in flight (Srinivasan, 1993). An ocellus also called an eyecup, is the simplest way to form a directional sensitive receptor. It is created by sinking a patch of receptor cells into the skin.

**Compound eye:** The compound eye is the most obvious and familiar visual organ of an insect. A compound eye is made up of a number of ommatidia. Each one is small, elongated eyecup with a crystalline cone at the tip and the light-sensitive photoreceptor below it. Every eyecup has its own lens, which sample an individual point in the visual field. At the moment, the assumption is that there is no overlap in the images caught by the individual eyecups. A transparent cuticle, the upper skin, covers the whole array of ommatidia (Bruce, Green, and Georgeson, 1996). Compound eyes vary in several ways around this basic plan. The number of ommatidia vary greatly. The eye of a blowfly consists of about 5000 ommatidia, which each have a field of view of about 1-2 degrees (Smakman, Hateren, and Stavenga, 1984). Thus the spatial resolution is much lower than in the vertebrate eye. Together the ommatidia sample an almost 360°-view of the surrounding visual environment.

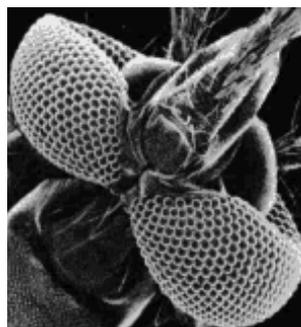


Figure 2.3: *The compound Eye of a Mosquito*

Another kind of variability is the degree of optical isolation between adjacent ommatidia. Depending on the task of the insect, the ommatidia ordering in the compound eye differs. Two researchers at the AI-lab investigated the evolution of the morphology of an artificial compound eye, in which the ordering of the ommatidia was adapted to create a balance between the morphology, neural control, task and the environment (Lichtensteiger and Eggenberger, 1999). They showed that for robots with different tasks the ideal placement of the ommatidia varied. This has also been found in different types of bees.

In comparison with the vertebrate eye, the spatial resolution of the compound eye is lower, but the temporal resolution is higher. A photoreceptor in the blowfly has an impulse response much faster than the impulse response in the human eye (Hateren, 1992). Therefore, the lower spatial and higher temporal resolution makes the fly's visual system less sensitive to motion blur (Schilstra, 1999).

For the purpose of flight control, navigation, prey capture, predator avoidance and mate finding, they obviously do a splendid job. Bees are able to memorize quite sophisticated shapes and patterns, and flies hunt down prey insects in extremely fast and acrobatic flight manoeuvres. Insects in general are exquisitely sensitive to image motion, which provides them with useful cues for obstacle avoidance and landing, but also for distance judgments.

### Gaze stabilization in a blowfly

Flies can see movement faster than humans can. The compound eyes of the blowfly are attached to its head and cannot move independently from the head. Thus head movements in flies are equal to eye movements. The head can move with respect to the thorax by means of an extensive system of muscles. During flight the head as well as the thorax make short, rapid turns about ten times a second. In humans we can see similar short rapid movements of the eye, called saccades, alternated with periods of stable gaze. Because of the resemblance, the fast turns in the blowfly are also called *saccades*.

In the horizontal plane, the thorax turns during the first 10 milliseconds while the head still looks in the original direction. The turning movements of the thorax are registered in a mechanical way by sensors called halteres, discussed above. After the movement of the thorax, the head rapidly turns in the same direction with a higher rotational velocity than the thorax and arrives earlier than the thorax (Schilstra, 1999). This head movement is comparable to the stabilizing head movements of a figure skater making a pirouette.

### The Consequences to the Visual System

Every visual system has an amount of inertia in processing the visual input. The first result is that fine details are blurred during simultaneous head movements and visual processing. Secondly, it results in disturbances in the normal optic flow. An eye and head movement adds noise to the optic flow, that later has to be separated from the normal optic flow by the visual system. The aim of the visual system is to keep the eye as stable as possible. The blowfly achieves this by moving its head faster than its thorax, minimizing the time-span of the saccade and maximizing the stable period.

### 2.2.2 Gaze Stabilization in Birds

One big difference between birds and mammals is that birds do not have only seven neck vertebrae or massive skulls anchoring powerful jaws. Instead there are many birds with small heads on long, flexible necks. Those birds are able to keep their visual input quite stationary with head stabilizing movements even during locomotion. The *gaze stabilization* experiments with birds are performed by measuring the head/eye movements of the bird while it walks on a treadmill. The bird does not fly during experiments.

Thus a pigeon or chicken *walking* does not suffer the visual streaming of the optic flow as mammals do, but keeps its head stationary with respect to the surroundings most of the time, while the legs move the body forward (Wallman and Letelier, 1989); when the legs have overtaken the head, the bird thrusts its head forward, a head saccade,



putting the head again in front of the body and then again stabilizing it with respect to the surroundings. Using this behavior the bird reduces interruptions of stable gaze by making fast head saccades, which is exactly the same behavior as the insect (explained in the section 2.2.1, page 8). During these head saccades, changes are also made in which direction the eyes are looking, thereby improving the stable visual input again (Pratt, 1982). The ability of

birds to stabilize their heads in space can be demonstrated by holding a long-necked bird in one's hand and twisting or translating its body; the degree of stabilization is so good that it almost seems as if the head is fixed in space with the neck passively joining

it to the moving body. In this case, gaze stabilization is achieved by head movements; the *vestibulo-collicular reflex*, neck movements based on vestibular signals.

In this situation, as in the walking one, many cues - proprioceptive, vestibular, and visual - help stabilize the head. A variety of experiments have demonstrated that *visual signals* are by far the most important, generally overriding conflicting signals from other sources. This visual predominance has been shown by several experiments on the head-bobbing during locomotion described just above. Friedman (1975) showed that when the visual environment is made to move with the head, walking pigeons do not bob with their heads anymore. Similarly, if the birds do not move with respect to the surroundings because they are on a treadmill head-bobbing also ceases. Even more convincing, if one positions a bird on a stationary perch surrounded by an oscillating visual environment, the head moves with the visual surroundings. This reflex is called the *opto-collicular reflex*, even though this visual stabilizing behavior perturbs the vestibular stabilizing mechanism.

In conclusion, the reason we mention the visual predominance is to alert the reader to the fact that visual information is the error-signal of the fine-tuning process of all the reflexes. And therefore the foundation of the gaze stability performance (Wallman and Letelier, 1989).

### 2.2.3 Gaze stabilization in Chameleons

Chameleons, animals that adapt the color of their skin according to their surroundings, generally live in shrubs and climb on small branches to catch small insects with their sticky tongues. To locate their prey, chameleons scan their environment with saccadic eye movements that are independent (non-conjugate) in the two eyes, although the two eyes are not able to track two targets simultaneously (Walls, 1942). This phenomenon is quite interesting; the chameleon is able to explore its surroundings twice as fast. Once the prey is located, the head is aligned and the two eyes converge to fixate the target binocularly (Gioanni, Bennis, and Sansonetti, 1993). The eye of the chameleon possesses a fovea, which is used in binocular vision. The eyes are placed laterally in tube-like enclosures which permit movements of 180 degrees horizontally and 80 degrees vertically (Walls, 1942). The chameleon's eyes are able to make completely independent large saccades that permit the animal to foveate targets with a single eye movement. This contrasts with birds and mammals, who often use multiple small saccades to acquire unusual targets. This allows the chameleon to scan most of its environment quickly. The assumption is this compensates for a limitation of the peripheral visual field due to the severely constrained view-angle of its tube-like pupil.



#### Optokinetic Reflex

As mentioned above, the eye movements are independent of each-other, i.e. hardly any binocular interaction between the two eyes, this also counts for the fast phases of the optokinetic reflex. Thus the chameleon uses binocular vision only during distance estimation of a located target. Another aspect of *gaze stabilization* worth mentioning is that during optokinetic stimulation, eye movements contribute more than head movements to *gaze stabilization*. In experiments in which the chameleon can move its head, most visual exploring movement are done by eye saccades (gain 0.5). If a head

movement occurs it is normally slow and smooth, but infrequent.

### Vestibulo-ocular Reflex

Vestibular reflexes were evoked by rotating the animal in the horizontal plane. When the animal was in the dark, the gain of the vestibular ocular reflex was very low (max 0.3). In the light and with the animal in a the head-restrained condition visuo-vestibular interaction improves the *gaze stabilization*. But in experiments in which the head was free to move and with visuo-vestibular stimulation (closest to natural conditions) *gaze stabilization* is optimal and presents a constant gain of 0.8 over the entire frequency range of stimulation studied (0.05-1.0 Hz) Thus, one can see that adding the visual stimulus to the experiment has an important increase in the gaze stabilization performance.

### Cervico-ocular Reflex

The cervico-ocular reflex (COR) stabilizes the eyes based on signals elicited by the neck muscles. In experiments this is evoked by restraining the head, while the body undergoes sinoidal movements. This stimulation provokes a compensatory cervico-ocular reflex (COR) with a gain of 0.2-0.4 as well as ocular saccades, which are especially numerous in the presence of a visual surrounding. The direction of these saccades in the chameleon occurs in the compensatory direction, i.e. opposite to the relative movement of the head with respect to the trunk. Additionally, the contribution of the COR in the actual gaze stabilization remains unclear (Gioanni, Bennis, and Sansonetti, 1993).

### Conclusion

Optimal *gaze stabilization* (gain of 0.8) is only obtained in chameleons with combined visual and vestibular stimulation in the free-head condition. This experimental setup approximates a natural situation. When the head is free to move, the chameleon uses both its head and its eyes to stabilize its gaze during horizontal motion of the optokinetic stimulus. Eye-head coupling in the form of concomitant fast head and eye movements appears to be exceptional in the chameleon as contrasted with most other vertebrates in which there is a tight linkage between fast eye and head movements (Collewijn, 1977). Overall one can conclude that the chameleon is making the fast eye saccades, while his head, if needed, makes smooth pursuit movements. Most amphibians, reptiles, and birds depend mainly on head reflexes to stabilize their gaze, while mammals use reflex movements of the eyes, their head reflexes being weak or even absent (McCollum and Boyle, 2001). Thus, the manner in which the chameleon stabilizes its gaze differs from that of other reptiles, and appears to be intermediate between that used by birds (see the pigeon example on page 11) and by mammals.

#### 2.2.4 Gaze stabilization in Rabbits

The interesting detail of the rabbit is that this animal lacks a fovea. Accordingly, its eye movements are fully dedicated to gaze stabilization. The oculo-motor system of the rabbit provides the opportunity to study *gaze stabilization* in a strictly isolated form. But the rabbit's eye does contain a horizontal zone of elevated density of ganglion cells and other receptive elements, the *visual streak*, that is aligned with the horizon (Tan, 1992). Due to this elongated receptive area on the retina, the rabbit enjoys almost panoramic vision. Since there is no distinctly preferred area in the projection of the horizon on the retina, the main concern of *gaze stabilization* in the horizontal plane is

that the surroundings are stationary projected on the retina. The moment the eye is almost disappearing in the eye-socket, the eye undergoes a fast resetting in the opposite direction, called *nystagmus*, to allow the rabbit to continue the tracking movement.

The input to the optokinetic reflex, the retinal image-slip, is conveyed by direction-selective ganglion cells in the retina (Tan, 1992). In the retina of the rabbit two types of direction-sensitive ganglion-cells were found. The first reacts as well as on the appearance of the light stimulus as to the disappearance, i.e. the on/off ganglion-cell. The second only reacts to the appearance of the stimulus, the 'on ganglion cell. These cells also show a preference in the direction in which they react maximally, in the opposite direction they do not show any activity at all.

Experiments in the rabbit showed that the optokinetic reflex in rabbits is most sensitive to very slow motion (1-10 grad of the retinal image). This was also shown in experiments with humans. This is in contrast to the vestibulo-ocular reflex that is most sensitive to fast head perturbations.

### 2.2.5 Gaze Stabilization in Primates

Primates have a frontal eye position, although many vertebrates have laterally placed eyes. Laterally placed eyes enable an animal to detect predators from any direction, but an advantage of a frontal eye position is that it increases the size of the binocular field, the segment of the view sampled by both eyes simultaneously. When the same information is obtained by two eyes, greater accuracy can be achieved in discriminating spatial and temporal patterns from noise. Information about distances of objects can also be obtained.

The frontal eye position must find a way to compensate for the ability to obtain information from any direction at one time. By rapidly changing the directional view by moving the head, eyes, or both, this constraint can be overcome. Most vertebrates can move their eyes to some extent, but few by large angles. Primates with a frontal eye position have all kinds of different eye movements (Bruce, Green, and Georgeson, 1996). In humans gaze movements are often made using the eyes alone (McCollum and Boyle, 2001).

Many studies, especially with humans indicate top-down influences like attention and cognition (Pola and Wyatt, 1993). Dubois found in 1978 that subjects with diminished attention had slower eye movements, but that only a small incitement was often enough to reach the maximum velocity again (Oey, 1979). Jung and Kornhuber (1975) noticed that the optokinetic reflex was dependent on the cooperation and attention of the subject. It appears that attention is important for providing a flexible response to the complex visual surrounding. The quality of the response is clearly dependent on the amount of attention of the subject, especially in tasks like tracking target motion across a complex background field.

Their eye movements can be classified into 3 categories:

#### 1. Gaze-shifting Movements

The purpose of gaze-shifting movements is to bring the image of a selected object onto the fovea and keep it there. The fovea, the high-resolution central area of the visual field, lies in the center of the retina.

**Saccades:** Rapid and intermittent jumps of eye position, that focus an object on the

fovea. As a person reads, the eyes make several saccades each second to scan the page. In humans, saccades are extremely rapid, often up to  $900^\circ$  per second.

**Pursuit:** Once an object is focussed on the fovea, pursuit movements keeps it on the fovea as it moves, or as the observer moves, at speeds below  $100^\circ$  per second (tracking).

**Vergence:** these movements adjust the eyes for viewing objects at varying depths. The object will be fixated by the fovea of both eyes. As an object moves closer, vergence movements will turn the direction of gaze of both eyes towards the nose. If an object comes too close, further vergence is impossible and diplopia (double vision) occurs.

## 2. Gaze-holding Movements

**Optokinetic-reflex (OKR):** This reflex is based on the retinal slip and corrects it. Retinal slip is defined as the overall velocity with which an image drifts on the retina (Shibata and Schaal, 2001). The goal of the OKR is to keep the image still on the fovea, which is the center of the retina, while making saccades to prevent the eye from moving in its corner. The reflex is slow (latency of 80-100 ms).

**Vestibulo-ocular-reflex (VOR):** The VOR achieves stabilization of the object in the visual field by controlling the eye muscles in such a way as to compensate for head movements. The VOR uses the head velocity signal acquired by the vestibular organ in the semicircular canals as sensory input. The VOR (latency of 15-30 ms) is faster than the OKR and can thus stabilize the image on the retina during rapid head movements (Shibata and Schaal, 2001). In normal primates the gain of the VOR, defined as eye speed over head speed, is very close to 1.0 even in darkness and at head speeds up to  $300 \text{ deg/s}$  due to its dependence on vestibular rather than visual stimuli (Burdess, 1996).

## 3. Fixational Movements

**Tremor:** The human eye is held in position by a dynamic balance between three pairs of muscles. Instability in this balance causes a continuous small-amplitude *tremor*.

**Drift:** Drift occurs during fixations and consists of slow drifts followed by very small saccades (micro-saccades) that have a drift-correcting function. These movements are involuntary.

## 2.3 Three-dimensional Projection of the Vestibular and Visual Signals

As mentioned before, the vestibular signals and the visual signals complement each other to achieve optimal *gaze stabilization*. To accomplish this cooperation between the VOR and the OKR, it is necessary for the brain to transform the different sensory input modalities of the two reflexes into a common coding in three-dimensional (3-D) space (Tan, 1992). The coding of the vestibular signals is directly related to the architecture of the vestibular canal system. The three semicircular canals are positioned orthogonally with respect to each other. By combining the output of the three canals, the brain is able to represent the speed of head rotation in three dimensions. (Burdess, 1996). This representation along three rotational axes has also been found in the optokinetic reflex. The input to the OKR, the retinal slip, is conveyed by direction-selective ganglion cells in the retina. By studying subsequent levels of processing of retinal-slip information it was recognized that flow-detection by ganglion cells and subsequent neural stages is organized in three pairs of channels, each pair

specialized in one of the three rotational axes, the same axes as those measured by the semicircular canals as well as those axes of the extra-ocular muscles (Wallman, 1993).

## 2.4 Characteristics of Optical Flow

Optical flow is the amount of motion that is projected on the retina. The goal as we have seen in the latter paragraphs is to minimize this optical flow on the retina as much as possible. Identifying an object while it flows over the retina is difficult. Try for instance to follow your moving finger while you try to identify an object in the background. As you notice there is a lot of irritating blurring. Optical flow has several characteristics.

### 2.4.1 Rotational Optical Flow

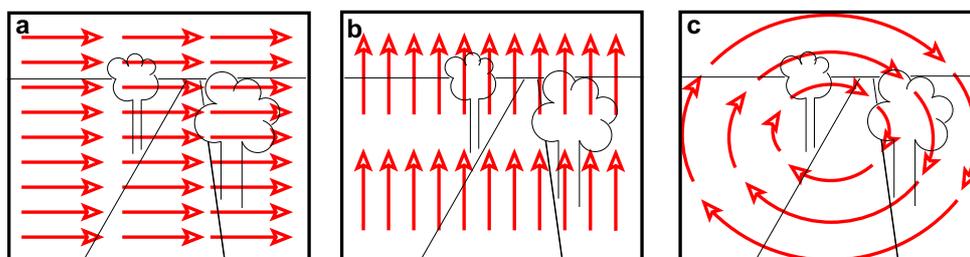


Figure 2.4: *Three types of rotational optic flow around the three axes. The rotation descriptions are taken from the insect movements. a* The 'yaw' movement, generates horizontal optic flow that is homogeneously spread over the image. For instance when the observer shakes his head. *b* The 'pitch' movement, in this case the vertical optic flow, is also homogeneous. An example is the observer nodding his head *c* In the 'roll' movement the optic flow is swirling around the point of focus in the image.

When the head rotates, the images of objects at different distances move coherently over the retina at approximately the same (rotational) velocity (Howard, 1993). The pattern and the speed of the optic flow at all points are determined entirely by the observer's motion - the 3D structure of the scene has no influence on the characteristics of the optic flow (Sandini, Panerai, and Miles, 2001). Rotational optic flow is elicited when the observer is rotating along a single axis. The head has three axes to rotate on, thus three types of optical flow can occur (see Fig. 2.4). The first is horizontal optic flow (*yaw movement*), elicited for example by a rotational movement of the observer moving his/her head from left to right. The second is vertical optic flow (*pitch*), which is elicited by an up- and down head-movement. The third is the roll movement which generates a circling optic flow. In reality pure optical flow like this never occurs. The axes around which the head turns are some distance behind the eye, so that the optic flow always has a slight translation, but in this thesis these effects were not significantly present in the image flow. This type of optic flow provides a source of information about the magnitude and direction of the eye or head movements (Rogers, 1993).

### 2.4.2 Translational Optical Flow

When the passive observer undergoes a pure translation, the optical flow has distinctively different characteristics. As shown in figure 2.5A the observer sees the optical flow emerging from its point of focus and speeding up as objects come nearer. In this case the 3D structure of the environment can be detected by this type of optical flow; nearby objects move faster on the retina. This phenomenon is called motion parallax (Sandini, Panerai, and Miles, 2001). If the observer is oriented side-ways the optical flow is spread over the retina according to figure 2.5B. Also in this case, objects further off move slower over the retina than objects closer by. Insects for instance use this feature to estimate the distance to an object (Srinivasan, 1993; Kootstra, 2002).

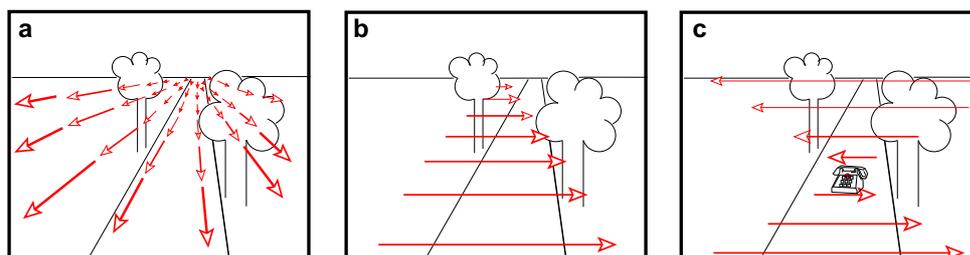


Figure 2.5: *Translational optic flow, a The characteristics of the translational optic flow, experienced by the observer looking in the direction of the heading. b The optic flow experienced by a moving observer looking sideways, without making stabilizing movements. c Again the moving observer is looking sideways, but this time the observer focuses on the telephone placed in the middle of the road. The scene now appears to pivot around this telephone.*

Humans have a symmetrical OKR and we can stabilize images of objects at one distance while ignoring distracting motion signals arising from objects at other distances, as we see in figure 2.5C.

## 2.5 What does Biological Information teach us in building Artificial Intelligent Systems

"There is clearly something we can learn from biology, . . . , we'd be foolish to ignore a billion years of evolution." (Prof. Dr. Dorio)<sup>2</sup>

Now that we know how a stable image is realized in biological systems, what does this information tells us? One thing for sure, natural and artificial systems that share the same environment may adopt similar solutions to cope with similar problems. Therefore biologists and engineers are often unintentionally focusing on the same topics, but from completely different points of view. The biologist analyzes behavior, while the engineer builds artificial systems with functionalities that fulfill certain desired behaviors. Occasionally, these behaviors overlap in functionality. In this case a dialogue between the two fields would be fruitful. The challenge is to find a common ground from where the biologist and the engineer can take advantage of a direct interaction.

The biologist can inspire the engineer, but on the other hand the engineer is able to give the biologist a better understanding of his described biological model, by building

<sup>2</sup>Prof. Dr. Dorio at the university of Washington

an artificial system according to the model. This approach is formally known as *understanding by building* and is discussed on page 4. Multidisciplinary research groups would be an ideal solution to stimulate the sharing of knowledge. The AI-lab in Zürich is an example.

Gaze stabilization is an example of common ground. The biological information discussed in this chapter, has provided a detailed overview of the gaze stabilization mechanism in natural organisms, in such a way that the engineer can create a formal model applicable for building an artificial system.

The biological information presented in this chapter gives us directly some ideas for building a gaze stabilization system in an autonomous agent. Why would there be two separate systems in biological systems to perform essentially the same image stabilizing task? The vestibulo-ocular reflex (VOR) is a very fast reflex with a latency of 15-30 ms (Shibata and Schaal, 2001) that serves mainly to compensate for the very fast head movements at frequencies in the range of approximately 1 to 7 Hz (Draper, 1998). However, the VOR is less accurate at lower frequencies, especially those lower than 0.1 Hz where the gain drops significantly and a phase lead appears. Over the range of frequencies contained in normal head movements, the gain of the VOR is only about 0.6 (Fuchs and Mustari, 1993).

The standard method in simulating a vestibular organ, is using a 3-axis gyroscope (Shibata and Schaal, 2001; Scassellati, 1998). The gyroscope is not sensitive to low frequency movements, in these two methods the optical flow is used to correct this problem. This project is planning to use slightly different hardware. To measure the tilt and roll, we want to use the ADX202E Dual-Axis Accelerometer with Duty Cycle Output. This hardware has been used for instance in autonomous flying machines (Stancliff, Laine, and Nechyba, 1994) to measure the tilt and roll. Their problem with the ADXL202 was that it is very sensitive to engine vibrations.

When the accelerometers oriented so both its X and Y axes are parallel to the earth's surface it can be used as a two axis tilt sensor with a roll and a tilt axis. Once the output signal from the accelerometer has been converted to an acceleration that varies between -1 g and + 1 g, the output tilt in degrees is calculated as follows:

$$Tilt = \arcsin(Ax/1g) \quad (2.1)$$

$$Roll = \arcsin(Ay/1g) \quad (2.2)$$

The accelerometer uses the force of gravity as an input vector to determine orientation of an object in space. It is most sensitive to tilt when its sensitive axis is perpendicular to the force of gravity. At this orientation its sensitivity to changes in tilt is highest. When the accelerometer is oriented on axis to gravity, i.e., near its +1g or -1g reading, the change in output acceleration is negligible. When the accelerometer is perpendicular to gravity, its output will change nearly 17.5 mg per degree of tilt, but at 45° it is changing only at 12.2 mg per degree and resolution declines. To measure the panning movements, we are planning to use an angular velocity sensor with a Murata ceramic bimorph vibrating unit from Gyrostar.

## 2.6 Conclusion

Compared to the VOR, the optokinetic reflex (OKR) has the opposite characteristics. It takes time to process the visual information, therefore a latency occurs of 80-100 ms (Shibata and Schaal, 2001). But at low frequencies, i.e. less than 0.1 Hz, the OKR has a gain of approximately 1.0 and no phase lead. From 0.1 Hz to 1.0 Hz the OKR loses gain and develops a phase lag due to its processing latency. Clearly, the vestibular and visual reflexes are complementary. Therefore the combination of the two mechanisms allow for maximal image stabilization across all frequencies.

On the other hand, we also realize that the retinal slip is used as an error signal for the tuning of the VOR, therefore a reliable implementation of the OKR has priority number one. In the next chapter we present an overview of some successful artificial gaze stabilization systems. After that we develop a biologically plausible model to detect optical flow, which will be tested in simulation and in a real-world office surrounding.

## Chapter 3

# Gaze stabilization in existing Artificial Systems

*Many artificial eye/head systems are developed in which gaze stabilization has been analyzed, since this mechanism is a crucial feature to address in building visually guided autonomous robots. Among other approaches in conventional vision- and robotic research, biological models have often been applied to artificial mechanical systems. Following are some examples of successful implementations;*

### 3.1 Shibata and Schaal

Shibata and Schaal developed a humanoid robot, able to stabilize its gaze by being capable of learning the accurate control of nonlinearities of the geometry of binocular vision as well as the possible nonlinearities of the oculomotor system. The control components in the oculo motor system resembled the known functional anatomy of the primate oculomotor system (Shibata and Schaal, 2001).

The optical flow was calculated by means of a block-matching method, i.e. the image is divided into a number of square blocks. The best-matching block compared to the template-image is found based on correlation by using least-prediction error, with minimization of the mean-square difference. This implementation is not really robust, except in the case of significant manual fine-tuning of the luminance differences and the image distortions. The learning system consisted of a biologically inspired feedback-error learning method combined with a non-parametric statistical learning network. By using eligibility traces, a concept from biology, and reinforcement learning, they solved the problem of the unknown delays in the sensory feedback pathways, for example delays caused by the complex visual information calculated. Shibata and Schaal (2001) actually created an accurate mathematical control plant similar to the human optomotor system.



## 3.2 The Babybot Project

This project, performed at the LiraLab, the university of Genova, focuses on the development of learning and adaptive behavior (Metta, Panerai, and Sandini, 2000). The acquisition of the appropriate behavior is obtained from an interplay with the agent's environment. The interaction of the agent's adaptive structure and the dynamics of its surrounding constrain the development of the otherwise difficult learning problem. The control structure of the robot evolves in stages therefore it constrains the learning process and potentially simplifies it. The agent is a physical robot interacting with the environment, i.e. the training set is collected on-line. Their interest is the development of the process of learning. The agent learns what it is capable of at that moment, constricted by the actual state of robot's system (Panerai, Metta, and Sandini, 2000). E.g. the robot cannot move the neck without controlling the eyes first. The robot is equipped with module containing three rotational and three linear accelerometers measuring head movements, similar to the vestibular organ in humans, plus an active vision module measuring the optical flow (Metta, 2000). The integration of these two modules results in a robust gaze stabilization system, that simulates the developmental stages similar to the stages a newborn child passes.

## 3.3 Cog Project

The goal of the Cog project at the MIT Artificial Intelligence Laboratory is to create a human-like robot that closely mimics the sensory and sensori-motor capabilities of the human visual system (Scassellati, 1998;2001). This robot should be able to detect stimuli relevant to humans, and respond to these stimuli in a human-like manner. Cog is capable of performing the following human-like eye movements:

1. saccades: fast eye movement towards an interesting object
2. smooth pursuit: maintains a (slow) moving object of the fovea
3. vergence: adjusts the eyes for viewing object at varying depths
4. vestibulo-ocular reflex: stabilize the eyes while the head moves based on a gyroscope
5. optokinetic reflex: stabilizes the eyes based on the retinal slip (image motion)

For the smooth pursuit eye movements, optical flow was measured as follows. A 7x7 pixels central window, the center of the camera image, was appointed as the target image. In the successive frames this windows was correlated with a 44x44 pixels central window. The 7x7 window was selected by an attention module based on three modules that searched in the image for motion, colorful objects and flesh-colored objects. The best correlation value gave the location of the target window in the new image, and the distance from the center of the visual field gave the motion vector. This motion vector was scaled by a constant and used as a velocity command to the motors. This resembles the method used in the Shibata en Schaal humanoid project in which they called this eye movements the optokinetic reflex. In the Cog project, another module calculated the optical flow of the entire background, which also resulted in a motion vector. In this case the optical flow estimate was the displacement vector for the entire scene and not only for this small 7x7 object-tracking window, although it was calculated in a similar manner.

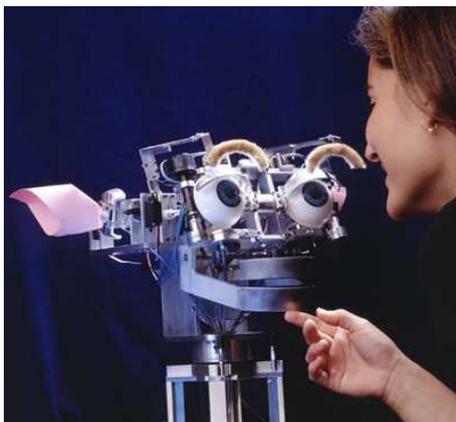


Figure 3.1: *The Humanoid Robot Kismet interacting with Cynthia Breazeal*

The vestibulo-ocular reflex was constructed by integrating the velocity signals from three rate gyroscopes mounted on orthogonal axes. Because of the integration, the systems tended to accumulate drift and the scaling constant to generate the appropriate motor command had to be selected empirically. This problem was solved with adding the optokinetic reflex. The optokinetic reflex (OKR) was used to train the vestibulo-ocular reflex (VOR) scale constant and the accumulated drift was corrected based on the optical slip.

### 3.4 The Kismet Project

We mention another a project at MIT, in this project social interaction with humans was investigated. Based on the capabilities of Cog, this project made another step towards a human-like robot. We mention this example to show that a humanoid robot equipped with a human-like gaze stabilization mechanism quickly gives humans the impression of social intelligence (Breazeal, 1999). The robot platform Kismet, a moveable head with a human-like face, consisted of an attention system that integrated separate perceptions, i.e. an auditory module, a motion detection module, a module sensitive to bright colors (toys) and a module able to recognize faces in the image. The goal of this attention system was to make the robot interact with humans. The interesting detail of this research was that the robot was not aware of the fact that it was communicating with humans, although it gave humans this feeling. Kismet was able to display a large variety of facial expression combined with head postures. In one experiment Kismet learned to react to people in an emotional way. By simply analyzing the change of the pitch in the human voice, the robot knew if he was praised or punished. The robot reacted accordingly with the appropriate facial expression and head posture. Its attention system also functioned in this experiment. The result was that the robot appeared to be socially intelligent (Breazeal, 1999). Another comparable approach in this field by (Kuniyoshi et al., 2000) has investigated engineering solutions for biologically plausible models of visual processing of a humanoid robot for versatile interaction with humans.



## Chapter 4

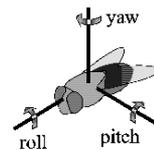
# A Biologically Plausible Model to Detect Image Motion

*We now know the characteristics of optical flow. How is this information calculated in biological systems? In the visual system of vertebrates on-off directionally selective ganglion cells have been found (Ibbotson, 2001) These cells respond much better to motion in one direction than to motion in the opposite direction. Aspects of the detection of image motion for gaze stabilization appear to be calculated by directionally selective ganglion cells (Amthor and Grzywacz, 1993). In this chapter we present the development of a motion detector based on the directionally selective ganglion cell, the elementary motion detector (EMD). The overall idea is to implement the optokinetic reflex (OKR), i.e., to measure the amount of optical flow in the image of the robot dog in order to stabilize its gaze.*

### 4.1 Processing of Optic Flow in Biological Systems

Directionally selective ganglion cells have been found in several vertebrates, although monkeys and cats do not possess many of them. In cats only about two percent of all ganglion cells are directionally selective (Amthor and Grzywacz, 1993). In rabbits and squirrels this percentage is 20 to 25 percent. The same amount is also found in amphibians, birds, turtles and reptiles. The response properties in all these species and also within species differ in sensitivity to movement of light compared to dark objects, preference for slow or fast speeds, and sensitivity to object shape.

In insects similar receptors have been found. Collett, Nalbach, and Wagner (1993) mention large-field neurons that sum the activities of many local motion detectors (LMDs), which give information about the direction of motion over a small area of the retina. These wide-neurons may be involved in decomposing optic flow (Collett, Nalbach, and Wagner, 1993). Two separate systems exist, one system sensitive to rotational flow and another system to translational flow. These systems makes use of large-field neurons, but how neurons are used depends on the type of flow the system detects (Wallman, 1993).



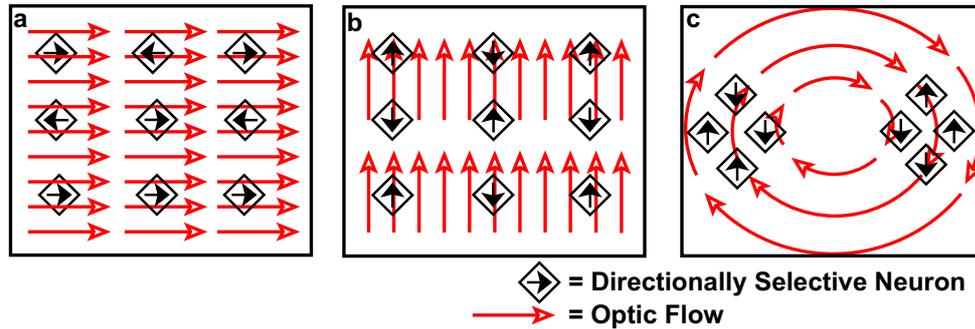


Figure 4.1: *The location of the Local Motion Detectors (LMD) in the retina. a Horizontal optic flow caused by a yaw movement, the LMDs are homogeneously spread over the image. b Vertical optic flow (pitch movement), again the LMDs are homogeneously placed. c Circling optic flow, caused by a roll movement, the LMDs are located where the optic flow is maximal.*

### 4.1.1 Directionally Selective Neuron

In the rotational system these neurons can be ordered in three groups; the first neuron type responds to yaw (fig. 4.1A), the second responds to pitch (fig. 4.1B) and the last is sensitive to roll (fig. 4.1C). Each type of neuron samples an array of LMDs with the appropriate preferred directions, so constructing a mask to pick up rotation about the desired axis. For instance, in case of a roll movement, a combination is made of the upward sensitive LMDs on one side with the downward sensitive LMDs on the other side (see figure)(Collett, Nalbach, and Wagner, 1993). A yaw or pitch movement results in an optic flow characteristic homogeneously spread over the image, and therefore measurable with two groups of LMDs (two directions) placed homogeneously across the image in the direction they are tuned for.

In the translational system large-field neurons are also combined to extract the translational optic information. It is difficult to tell whether a neuron is used by the rotational or the translational system (Wallman, 1993). One assumption about the development of this translational system is that the vestibular system affects the connectivity of the visual system neurons so as to create a visual analog of each of the semicircular canals. The visual system learns which neurons to examine to extract the right type of optic flow (Wallman, 1993). This might be an interesting topic for future research.

## 4.2 Development of the Elementary Motion Detector

The elementary motion detector (EMD) model was proposed for the calculation of image motion and is based on experimental data, obtained from optomotor research in insects. This model has functions which are similar to the directional selective ganglion cell. Evidence of EMDs in vertebrates has been noted by Borst and Egelhaaf (1989).

### 4.2.1 Criteria for Calculating a Directional Response

Several criteria must be met in order to calculate a directional response from a moving image (Borst and Egelhaaf, 1989). First, two spatially displaced inputs are necessary to model motion as a vector. Second, there must be temporal asymmetry in the processing of the signals from the two inputs. Third, the interaction between the signals must be non-linear. The non-linearity is essential because in a purely linear system the mean time-averaged output would be the same for opposite directions of motions.

According to these criteria, Reichardt developed the elementary motion detector (EMD) model in 1956, based on the optomotor response in insects (Reichardt, 1969). The non-linear factor in this model is multiplication, and the asymmetry is represented by delaying one input signal, which will be thoroughly treated in the next section. Several derivatives of his models have been developed based on his findings. The motion-energy model by Adelson and Bergen (1985) and the Barlow-Levick Model by Barlow and Levick (1965) are two examples.

The EMD model has been implemented successfully in several robotic projects. For example, Viollet and Franceschini (1999) built an aerial minirobot which stabilized in *yaw* by means of a visual feedback loop based on a single EMD. Iida (2000)(2001) successfully implemented a biologically inspired model of the bee’s visual “odometer” based on EMDs. The model was used to estimate the distance traveled based on the accumulated amount of optic flow measured by EMDs. The model was implemented on an autonomous flying zeppelin robot, called Melissa. Kootstra (2002) also using Melissa, created a landmark selection system based on EMDs. By using the characteristics of the translational optic flow, i.e. objects closer by move faster on the retina, relevant nearby landmarks for visual navigation were selected.

### 4.2.2 The Half-Detector

In 1956 Reichardt came up with his elementary motion detector (EMD) model. To give a clear picture of what this EMD is capable of, first we will introduce the half-detector (see Fig. 4.2). The half-detector is directionally selective, similar to the functionality of the Local Motion Detector (section 4.1). After this section we will present the elementary motion detector itself, which consists of two anti-symmetrical half-detectors, and thus able to detect optic flow in two directions.

The detector consists of two photoreceptors ( $P_1$  and  $P_2$ ). Each photoreceptor has three input values per pixel; red, green and blue. It gives as output the luminance, calculated as follows;

$$L = (R + G + B)/3.0 \quad (4.1)$$

Furthermore, the half-detector has two high-pass filters ( $H_1$  and  $H_2$ ), a low-pass filter ( $L_1$ ) and a multiplication unit ( $M_1$ ). This half-detector extracts a motion signal from the spatiotemporal correlations that are present in the moving image in only one direction (Reichardt, 1969). Let’s take a closer look at fig. 4.2.

#### High-pass Filter ( $H_1$ & $H_2$ )

$$H_1(t) = P_1(t) - P_1(t - 1) \quad (4.2)$$

The goal of the filter is edge-detection. The input signal  $P_1(t)$  contains only luminance information. As you can see in figure 4.2, the image that moves across the photoreceptors, is first black, then white then black again. The filter only reacts to changes in luminance.  $P_1(t)$  and  $P_2(t)$  contain the luminance information that are received by two photoreceptors separated by distance  $S$ . By subtracting  $P_1(t)$  from  $P_1(t - 1)$  this filter only transmits the signal change.

#### Low-pass Filter ( $L_1$ )

$$L_1(t) = \alpha H_1(t) + (1.0 - \alpha)H_1(t - 1) \quad (4.3)$$

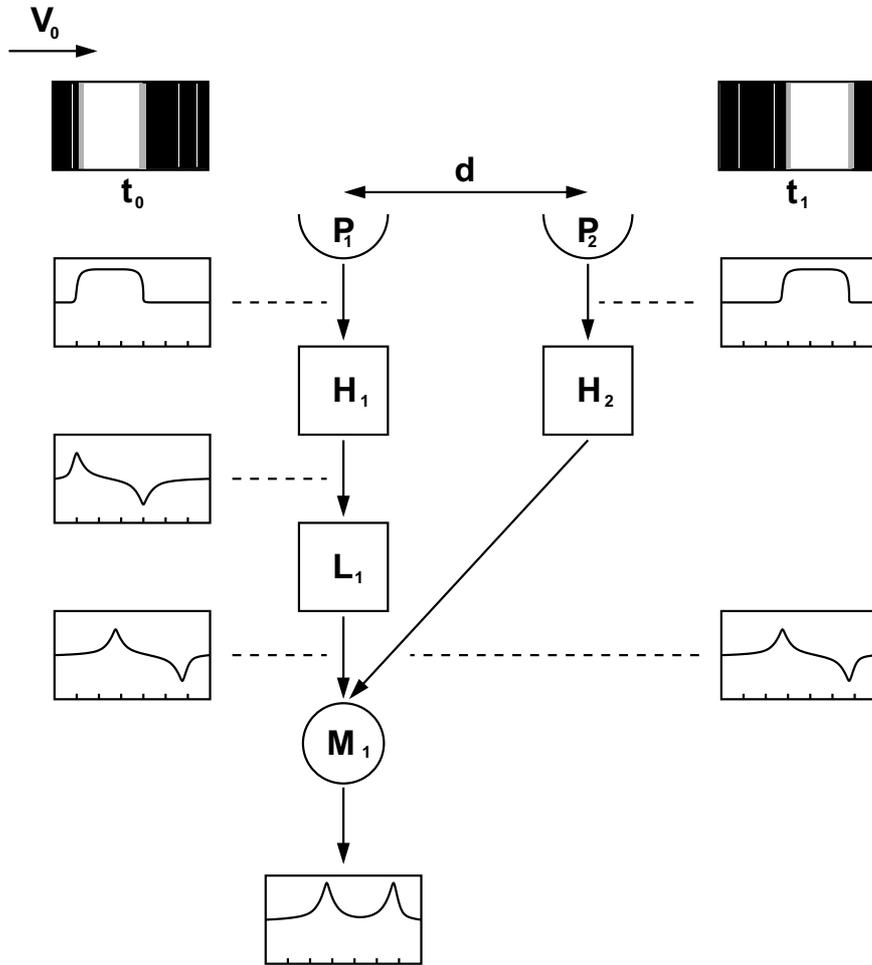


Figure 4.2: *Half-Detector, a one-directional motion detector.* The half-detector perceives luminance information from the image. The detector consists of two photoreceptors ( $P_1$  and  $P_2$ ), two high-pass filters ( $H_1$  and  $H_2$ ), one low-pass filter ( $L_1$ ) and a multiplication unit ( $M_1$ ). The photoreceptors are separated from one-another by distance  $d$ . The stimulus moves from left to right with velocity  $V_0$  across the photoreceptors.  $V_0$  is the velocity this half-detector is tuned for.  $V_0$  depends on; 1.  $(t_0 - t_1)$ , 2. how fast the perceived edge crosses  $d$  (the distance between the photoreceptors) and 3.  $\alpha$  (the parameter of the low-pass filter). This figure shows a perfect match situation, the stimulus moves exactly as fast as the low-pass filter delays the signal, and therefore the signals  $L_1$  and  $H_2$  cause  $M_1$  to give a high correlation signal (two peaks = two matched edges). The full EMD contains an additional mirrored version of the half-detector, and the detector becomes bi-directionally sensitive

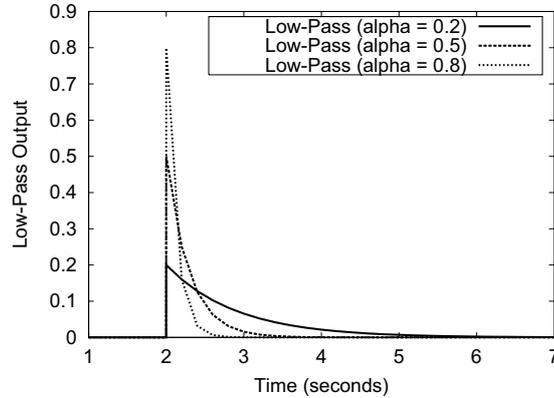


Figure 4.3: *The characteristics of the Low-Pass filter, in this figure  $\alpha$  was varied. As you can see the smaller the  $\alpha$  the longer and smaller the response from the low-pass filter.*

The low-pass filter is a time delay filter, the smaller the value of  $\alpha$  the bigger the delay. The parameter  $\alpha$  specifies the image speed this half-detector is tuned for. Actually the graph in figure 4.2 is incorrect, in this figure we used a real time delay. If you take a closer look at equation 4.3, you will see that the signal should slowly decay to zero after receiving an impulse. This decay is set by parameter  $\alpha$ . According to eq. 4.3  $L_1(t)$  is  $(1.0 - \alpha)$  times the previous value plus  $\alpha$  times the signal delivered by  $P_1(t)$ .

#### Non-linear Filter ( $M_1$ )

$$M_1(t) = L_1(t)H_2(t) \quad (4.4)$$

This filter compares two signals, in this case  $L_1(t)$  and  $H_2(t)$ . Multiplying these signals results in a peak if the time-delay is exactly as long as it takes for the edge to move distance  $d$  across the photoreceptors. Otherwise no match is found and  $M_1(t)$  will not fire. We have now constructed a directionally selective motion detector, i.e. it reacts only to stimuli in its “preferred direction” (see fig. 4.4) and does not give a response for motion stimuli in its “null-direction”.

#### 4.2.3 The Elementary Motion Detector (Reichardt Detector)

An elementary motion detector (EMD) consists of two mirror-symmetrical half-detectors. The “preferred direction” in one half-detector is the “null-direction” and vice versa in the other. The only addition is the subtraction unit. The response from the second half-detector shows exactly the same response characteristics to motion in the opposite direction as the half-detector in figure 4.2.

#### Role of the Subtraction Unit ( $S$ )

$$S(t) = M_1(t) - M_2(t) \quad (4.5)$$

The signals from the two half-detectors have the same polarity, therefore the subtraction unit is added to the model.  $M_1(t)$  detects image motion from left to right, and  $M_2(t)$  has the opposite preferred direction, thus the polarities of the two signals are

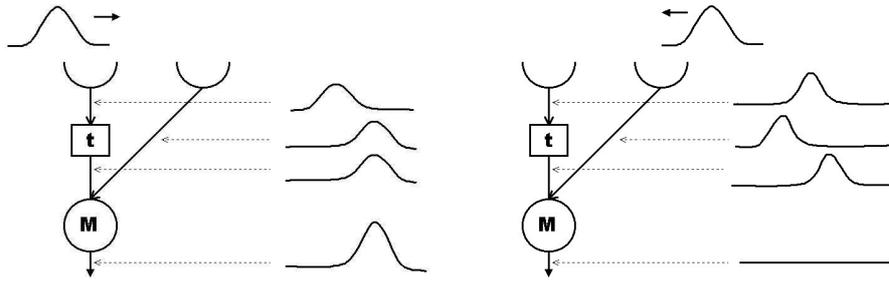


Figure 4.4: For motion in one direction (“the preferred direction”) the time-delay filter compensates for the time-shift caused by the photoreceptor distance and the half-detector fires. For motion in the opposite direction (“the null-direction”) the time-delay causes that the multiplication unit finds no match at all. Note that in this figure  $t$  is an absolute time-delay. We use this type of delay to explain the functionality of the EMD. In our implementation we used a low-pass filter  $L_{1,2}$  in fig. 4.2

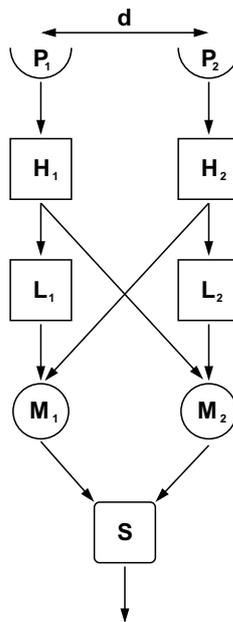


Figure 4.5: A full Elementary Motion Detector, consisting of two photoreceptors ( $P_1$  and  $P_2$ ), two high-pass filters ( $H_1$  and  $H_2$ ), two low-pass filters ( $L_1$  and  $L_2$ ), two multiplication units ( $M_1$  and  $M_2$ ) and a summation unit ( $S$ ). This model is a bidirectional selective motion detector.

also contrary. Otherwise it is not possible to distinguish the two opposite directions of the optic flow across the photoreceptors. This results in a bidirectional selective motion detector.

$$\begin{aligned} \text{if } M_1 < M_2 &\Rightarrow S < 0 && \text{Motion from left to right} \\ \text{if } M_1 > M_2 &\Rightarrow S > 0 && \text{Motion from right to left} \end{aligned}$$

As a general reference to the EMD function of the two pixel luminances we will use:

$$E_n(t) = \Gamma(P_1(t), P_2(t)) \quad (4.6)$$

### 4.3 Implementation of the Motion Detector

We will now test the characteristics of the current motion detector model. We are looking for a motion detector that gives a deterministic, linear reaction to the perceived image velocities. The ideal curve would be a linear curve that gives exactly the same response for each bar width. Thus in the ideal case we demand a detector that is not dependent on the characteristics of the surroundings (in this case the bar-width) and only depends on the amount of the image flow.

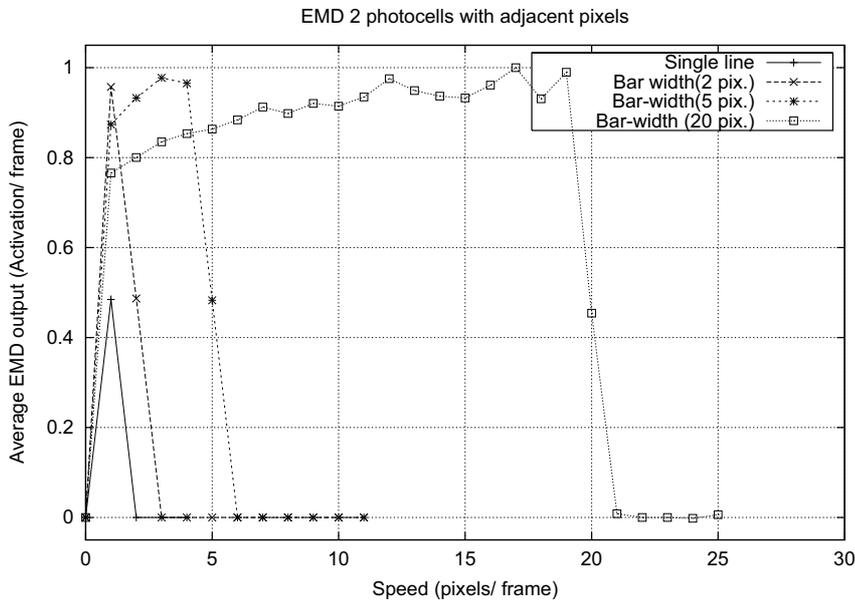


Figure 4.6: In this graph we present the performance of the elementary motion detector (EMD) with adjacent photoreceptors for four moving patterns. The results are obtained in a simulated environment in which for every experiment bars with different widths and different speeds moved across the EMDs on a black background. The y-axis displays the average EMD-output per frame, and the x-axis displays the speed in pixels per frame. Note that, based on the average EMD output, no distinction between different image-flow velocities can be made, therefore this model is not good enough. We want to see a linear relation between EMD-output and speed.

#### 4.3.1 EMD Results in Simulation

The elementary motion detector is sensitive to image motion. In this section we present the characteristics of the EMD with adjacent photoreceptors in a simulated

environment. The simulation consists of a simple black image in which one hundred EMDs were placed homogeneously in the image. Across this black background a white bar, with four different bar-widths (1, 2, 5 and 20 pixels) moves with an increased speed per cycle of the experiment. Speed is defined as the amount of pixels by which the bar-position is changed for every frame. In fig. 4.6 we plotted the outcome of this experiment. The y-axis displays the average EMD output per frame and the x-axis displays the speed of the respective white-bar per frame. The average EMD output per frame is calculated by averaging the total EMD output sum of the hundred EMDs in the image examined per frame. As we can see in fig. 4.6 the EMD mostly reacts maximally to image motion, and is therefore only sensitive to one particular value of image flow velocity, i.e. based on the average EMD output no distinction between different image flow speeds can be made. As we see, a standard EMD approach is able to detect the image flow velocity  $v_0$  it is tuned for. But it is not able to give a quantitative judgment of image flow velocity. Therefore, in the next section an improved version will be proposed.

### 4.3.2 Improvement of the Motion Detector

With the composition of the EMD, a motion detector is created that reacts maximally to its optimal velocity  $v_0$ . This velocity,  $v_0$ , can be varied in two ways:

- Changing the distance  $d$  between the photoreceptors
- Changing the time-delay parameter  $\alpha$

Changing the time-delay parameter  $\alpha$  has two disadvantages. First, if the parameter setting is specified for detecting slow image motion, a small  $\alpha$  will result in a motion detector tuned for a small  $v_0$  image velocity. But the slope of the decay causes that the low-pass filter to send a (decaying) signal (see fig. 4.3). This results in the fact that other faster edges will also react with the signal from the low-pass filter. Therefore you cannot be sure if the correlation found by the EMD results from the same edge.

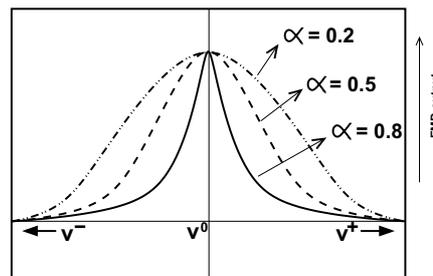


Figure 4.7: This graph displays the dependency of the EMD sensitivity ( $v_0$ ) on the low-pass parameter. The  $\alpha$ -parameter tunes the sensitivity to the image velocity. The smaller the  $\alpha$  the more sensitive the EMD is to a wider range of image flow.

On the other hand, again in case of a small  $\alpha$  other edges will enter this low-pass filter, before it empties itself after the first edge. Therefore the output-signal of this low-pass filter will increase although the edges do not move faster. This will result in a higher response from the EMD. This means that the EMD becomes more and more dependent on the amount of edges, in other words dependent on the characteristics of the surroundings. An EMD detects if its  $v_0$  velocity is present in the image motion.

But the accuracy depends on the parameter  $\alpha$ . In case of a small  $\alpha$  the EMD responds to a large variety of image motion (see fig.4.7), but this is not what we are looking for, because in this case we lose detailed information about the exact amount of image flow.

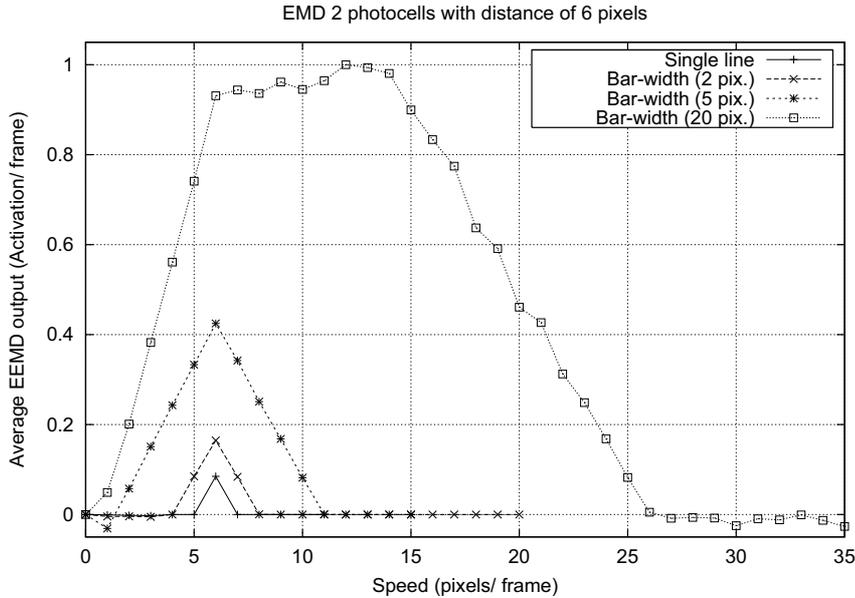


Figure 4.8: In this graph we present the performance of the elementary motion detector (EMD) with a distance between its photoreceptors of five pixels for four moving bar-widths. The results are obtained in a simulated environment in which on a black background bars with different widths and different speeds per experiment moved across the EMDs. The y-axis displays the normalized average EMD-output per frame, and the x-axis displays the speed in pixels per frame. Note the distance of 5 pixels between the photoreceptors. This causes the EMD to show a maximum response for image velocity  $v_0 = 6$  pixels per frame. Thus by alternating the distance between the photoreceptors we can construct an EMD tuned for a particular image speed  $v_0$ . But still this is not good enough, the EMD does not react anymore to small objects.

The final argument against changing the parameter  $\alpha$  to create EMDs that are tuned for different velocities is, that with changing  $\alpha$  slower image motion becomes measurable, but in case of gaze stabilization motion detection also includes fast image motion. The frame-rate of our artificial system was fixed at 20 Hz. Detecting faster image motion is a prerequisite.

A single EMD is capable of detecting whether the image motion approximates the  $v_0$  velocity of the EMD. The goal of a gaze stabilization system is to adjust the eye velocity to the image velocity, so that there is no more retinal slip. Therefore we need to construct a motion detector able to provide a quantitative signal of the image motion. As displayed in fig. 4.8 we see that changing the distance between the photoreceptors results in this altered EMD firing maximally at another velocity  $v_0$  compared to the

EMD with adjacent photoreceptors.

#### 4.4 The Extended Elementary Motion Detector

The solution is simple, we present the extended elementary motion detector (EEMD). This EEMD consists of several EMDs each with a different distance  $d$  between its photoreceptors (Fig. 4.9). Therefore an EEMD is sensitive for several image-flow velocities, each EMD is tuned for a different  $v_0$ .

$$E_1(t) = \Gamma(P_3(t), P_4(t)) \quad (4.7)$$

$$E_2(t) = \Gamma(P_2(t), P_5(t)) \quad (4.8)$$

$$E_3(t) = \Gamma(P_1(t), P_6(t)) \quad (4.9)$$

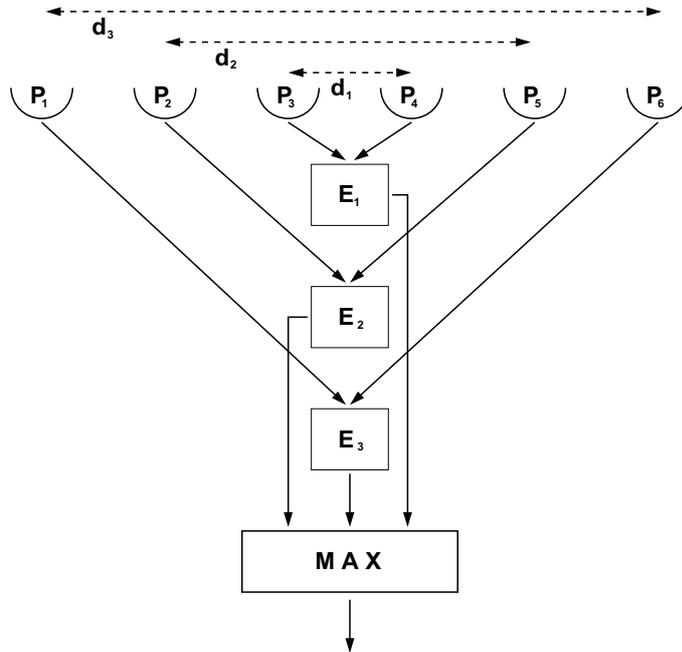


Figure 4.9: *Extended Elementary Motion Detector (EEMD)*.  $E_1$ ,  $E_2$ ,  $E_3$  are elementary motion detectors (with the same low-pass parameter  $\alpha$ ). Each EMD has its own two photoreceptors with a different  $d_n$  between its photoreceptors. Each EMD is therefore tuned for a different image motion  $v_0$ . The weighted output of each EMD is analyzed by the MAX unit, in which a winner-takes-all principle prefers the EMD with the highest value and multiplies this signal with  $d_n$ , the distance between the photoreceptors of the particular 'winning' EMD.

where the  $\Gamma$ -function stands for the EMD calculation (eq. 4.1 - 4.5), presented in equation 4.6 .

where  $E_n(t)$  is the bidirectional motion signal detected by the EMD  $E_n(t)$ .  $P_n(t)$  are photoreceptors that receive luminance information.

$$d_{max} = d_{arg \max_{n=1}^3 (E_n(t))} \quad (4.10)$$

$$EEMD(t) = d_{max} \max_{n=1}^3 (E_n(t)) \quad (4.11)$$

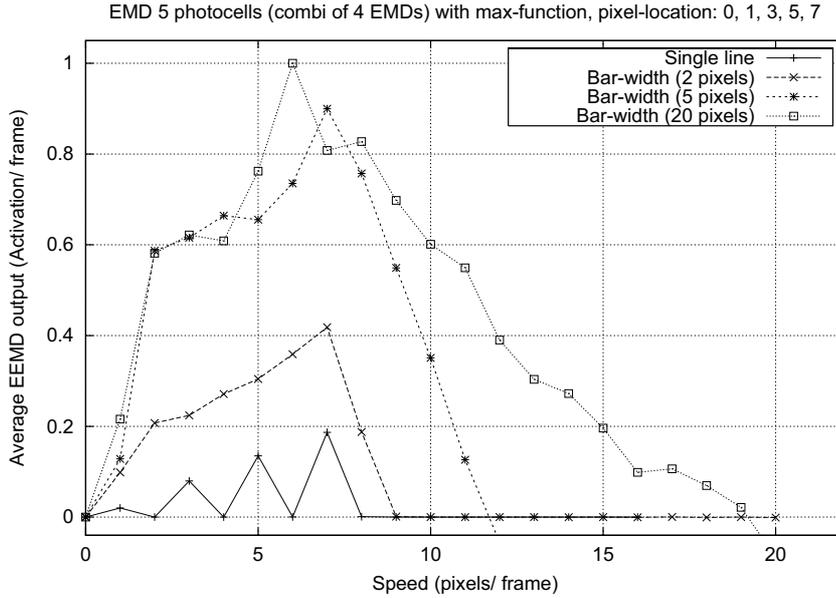


Figure 4.10: This graph represents the performance of the extended elementary motion detector (EEMD). The EEMD consists of 4 EMDs, with distances between the photoreceptors of 1, 3, 5 and 7 pixels. The results are obtained in the same simulated environment as in fig. 4.6 and 4.8. The y-axis displays the average EEMD-output per frame, and the x-axis displays the speed in pixels per frame. Note that the response to a single line clearly shows the four EMDs with their own response multiplied by  $d_{max}$  from equation 4.10. This graph already comes closer to the graph that we are looking for, because every curve contains an aspect of a deterministic relation.

#### 4.4.1 EEMD Results in Simulation

By constructing the extended elementary motion detector (EEMD), we have created a detector able to detect four different image velocities. We have tested this EEMD in the same simulated environment, i.e. a moving whitebar on a black background with bar widths ranging from one to twenty pixels, as used in testing the EMD characteristics. Speed is defined as the amount of pixels that the bar-position is changed every frame. In graph 4.10 the characteristics of the EEMD output are plotted. The x-axis represents the speed of the respective white bar in every frame. The y-axis displays the average EEMD output. Again one hundred EEMDs are homogeneously placed across the image and the total EEMD output was divided by the number of EEMD's (one hundred). The curve of the "single line"-bar clearly displays the four separate EMD

responses multiplied by the distance between the corresponding photoreceptors. The curve of the bar with a width of two pixels results in a continuous version, because the low-pass filter value is not empty yet. The graph 4.10 does not show us exactly what we are looking for but the curve does show a deterministic response for every bar width, although not perfectly linear.

## 4.5 Discussion

As we see, the response for every bar-width quite different. The width of the white bar influences the curvature. In a real-world implementation we do not expect difficulties with this feature. Our assumption is that in a real-world environment all bar widths (or any other variations) are present simultaneously through a wide variety of object, colors etc. Therefore it is hypothesized that placing several EEMDs in a dispersed order in the camera image and averaging the total EEMD output per time cycle will result in a smooth deterministic graph of the detected image motion. In the next chapter we will test the latter extended elementary motion detector (EEMD) in a real-world office environment and apply this model to obtain gaze stabilization in the robot dog.

## Chapter 5

# Gaze Stabilization in the Robot Dog

*By means of the extended elementary motion detector (EEMD), which calculates the optical flow, we will use the motion signal to stabilize the gaze of the robot dog. Thus gaze stabilization is based on visual information, and is therefore analogous to the optokinetic reflex in biological systems. The goal of this project is to stabilize the gaze of the robot dog for rotational optic flow elicited by ego-motion. This means, practically, the dog nods its head up and down to causes vertical optic flow (fig. 2.4B) or the dog shakes its head from left to right, to causes horizontal optic flow (fig. 2.4A). While its head moves, the eyes perform compensating eye-movements. Before we present the gaze stabilization result, let's take a closer look at the robot dog itself.*

### 5.1 The Robot Dog

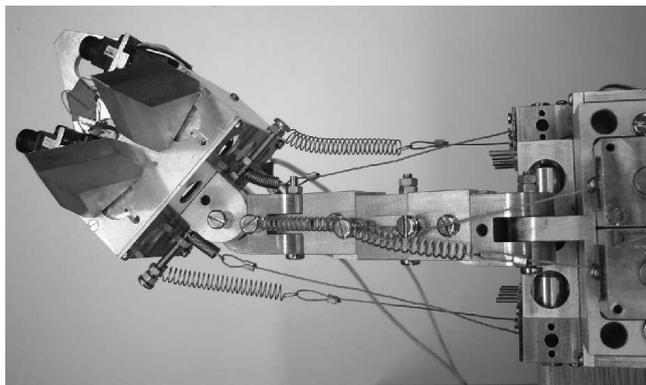


Figure 5.1: *The musculo-skeletal design of the neck/head system (top view). The head (on the left side) and the neck are connected to the motors in the front part of the body (on the right) with artificial 'tendons' (wires and springs). These tendons are fixed to a winding axle of a servo-motor*

Inspiration by biology is the key-factor in the design process of the robot dog. We have thoroughly analyzed the anatomy of the biological dog (Goody, 1997). This natural system has evolved and optimized its adequate functionalities through interaction with

its environment. Therefore we carefully considered the physical properties of the robot dog, such as its sensors, actuators, materials and body-structure. Recent studies in embodied cognitive science show that an appropriate design of the morphology helps the learning process significantly in successfully fulfilling a certain task. In the design of the robot dog's morphology we tried to imitate the physical properties of a natural dog, i.e., the proportion and weight of the skeleton structure, the number and position of the joints and the location of the actuators.

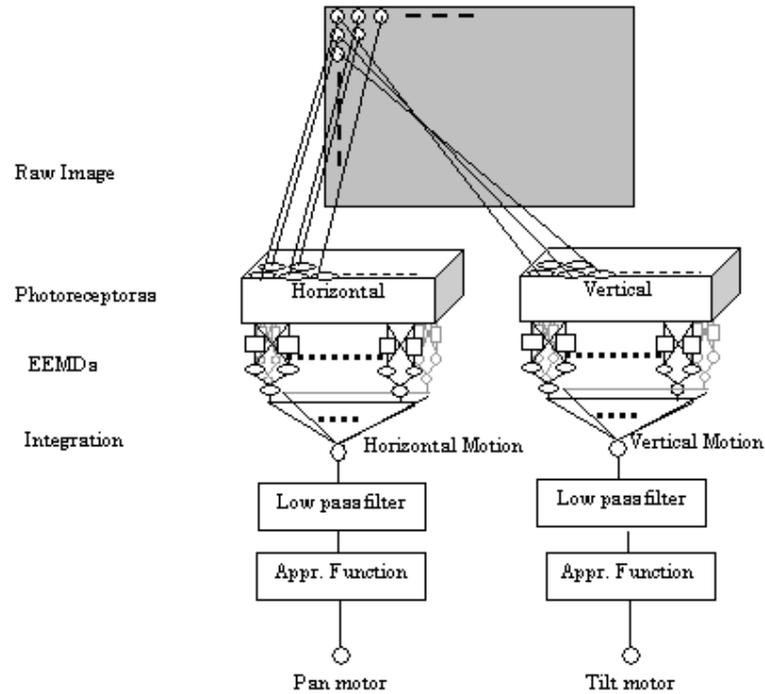
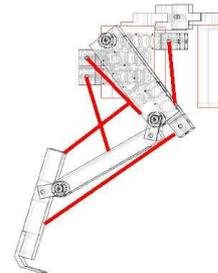


Figure 5.2: *The control architecture of a pan-tilt active vision system. Motion detectors cover the entire image along both the horizontal and vertical axes. The outputs are spatially integrated and averaged by the number of active EEMDs, and finally low-pass filtered. The corresponding motor outputs are determined by using function approximation based on the EEMD-output.*

The skeleton of the robot is constructed out of aluminium, and the dimension of the body is approximately 750x600x300mm. The skeleton contains 28 passive joints, the rotational angles of each joint are mechanically restricted, and each joint is capable of a small translational displacement as well. The passive joints are controlled by a muscle-like actuation method in which we incorporated electric motors connected with a wire, and a spring in series. This *artificial muscle* mimics the properties of the muscle-tendon system of animals. Thus one motor can control multiple passive joints. This approach has the advantage that



joints are controllable in a non-linear fashion, while it closely mimics biological systems.

In the following section we focus only on the eye/head/neck part of the robot for the gaze stabilization task. The six muscles are installed in the frontal part of the body which control eight passive degrees of freedom in the spinal block of the neck (fig. 5.1). In the skull of the robot a binocular active vision system is placed, which consists of four servo-motors (S3106 Futaba) and two miniature CMOS cameras. These servo-motors in the active vision system have a maximum angular speed of approximately 550 deg/sec. The maximum rotational angles are restricted to 120 degrees in the horizontal direction and 65 degrees vertically. All the motors are controlled by an external host-computer through communication with micro-controllers. Visual processing is also conducted at the same host-computer in which two frame-grabber devices are installed.

## 5.2 Experimental Results

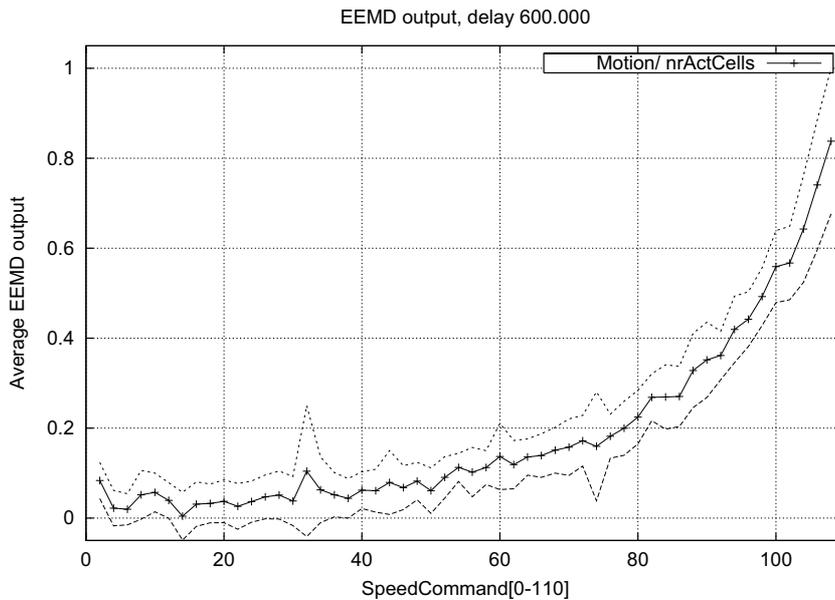


Figure 5.3: This graph represents the EEMD output during a motor velocity gradually increasing from 0 to 110. We obtained this graph by sending commands ranging from 0-110 to the motors while recording the EEMD-output. We repeated the experiment 100 times, and averaged the outcome. Additionally the standard deviation of the EEMD-output ( $\pm\sigma$ ) is plotted as the dashed lines. The delay of 600.000, mentioned in the header of the graph, is the delay with which commands were sent to the motor board to control the eye motors. This graph shows that the hypothesis assumed in chapter four is correct. The relation between speed and EEMD-output is now deterministic, and usable for gaze stabilization.

Concluding from figure 5.3 we clearly see that the extended elementary motion detector (EEMD) shows a beautiful deterministic response, which is usable for gaze stabilization. The hypothesis proposed in the previous chapter on page 4.5 is confirmed. This

graph is obtained in a real-world office environment and shows that our assumption that the characteristics of this environment have a smoothing effect on the EEMD-output is correct.

In our experiments we implemented 3200 vertical and 3120 horizontal EEMDs. The reason we used this number is that, while we performed the experiments in a real-office environment, a lot of noise was present. The chance of a single EEMD giving an erroneous output is quite high and averaging over several EEMDs minimizes this chance. The EEMDs cover 69x92 degrees in absolute angles, which is digitized into a 160x120 pixel image. For the gaze control, the outputs of the EEMDs are spatially integrated into mean vertical and a mean horizontal speeds, which are proportionally connected to the eye servo-motors after low-pass filtering of the EEMD-output (fig. 5.2). The motors are then controlled in such a way that apparent image motion induced by ego-motion is canceled. Since the output of the proposed detector is highly non-linear (fig. 5.3), we need to find the correct mathematical function to translate the motion signal to the appropriate motor command, so that it compensates for the perceived amount of motion.

### 5.2.1 Function Fitting

In order to achieve live gaze stabilization, we have to find a function that matches the Extended Elementary Motion curve presented in figure 5.3. This graph is created by sending motor commands ranging from 0 to 110 to the eye motor, while the EEMD module measured the amount of motion. As we take notice of the graph, it shows a non-linear curve. On the other hand the characteristics of the motor are also non-linear, as we see in fig. B.2 in appendix B. In this graph we plotted the given motor command against the actual motor speed in degrees per second. What we are looking for, is the inverse curve of the EEMD plot (fig 5.3) Two methods were used: first, a mathematical approach in which the formal exponential curve was approximated and second, a neural network approach, in which we let a neural network learn the relationship between the EEMD output and the corresponding motor command.

### 5.2.2 Exponential Curve Fitting

The main focus in programming the robot dog is minimizing the computational load, while the whole system operates real-time. Therefore speed-optimization is a very relevant aspect. We first tried a piece-wise linear approach, because it is simple and fast.

The idea is as follows, we approximate the EEMD-curve with two linear equations shown in figure 5.4. The appropriate values of a, b, c and d were deducted from the EEMD-curve. But we discovered that the EEMD-output curve was non-linear and therefore this piece-wise linear approach was not good enough.

Examining figure 5.3, we notice that the form of the EEMD-curve shows the typical form of an exponential function.

$$f(x) = a * e^{b(x-c)} + d \quad (5.1)$$

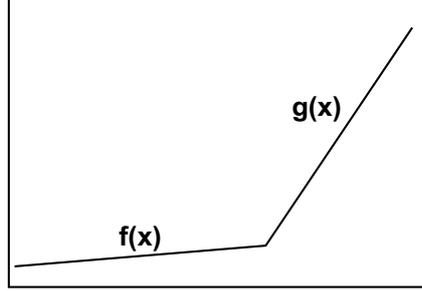


Figure 5.4: This graph presents the form of the two linear curves that are used to approximate the EEMD-output function (for details, see text)

By using a mathematical approach we can construct an exact replication of the EEMD curve as follows:

- a Maximum value of the EEMD-output
- b Roundness of the exponential curve
- c Maximum speed measurable by EEMD
- d Vertical shifting of the graph, the minimum EEMD value

Parameters a and c were deduced manually from the graph in fig 5.3. Parameter b represents the roundness of the exponential curve, calculated as follows:

$$f(x) = y = a * e^{b(x-c)} + d \quad (5.2)$$

$$\left(\frac{y}{a} - d\right) = e^{b(x-c)} \quad (5.3)$$

$$\ln\left(\frac{y}{a} - d\right) = b(x - c) \quad (5.4)$$

$$b = \frac{\ln\left(\frac{y}{a} - d\right)}{(x - c)} \quad (5.5)$$

We acquire the correct value for b by filling in the EEMD-value that is elicited with a motor command of 60 to approximate the roundness of the exponential function. In the implementation we used the inverse-function of the exponential curve:

$$speed = \frac{\left(\ln\left(\frac{EEMD}{a} - d\right) + bc\right)}{b} \quad (5.6)$$

### Results Exponential Curve-Fitting Method

To test the performance of this approach, we let the motor swing from left to right with different motor commands and we let the function approximation module estimate what motor command was given to the motor based on the visual signal resulting from the EEMD module. The idea behind it is, if we can estimate the given motor command based on the measured image flow, we can also generate the same motor command, which will result in stabilizing behavior.

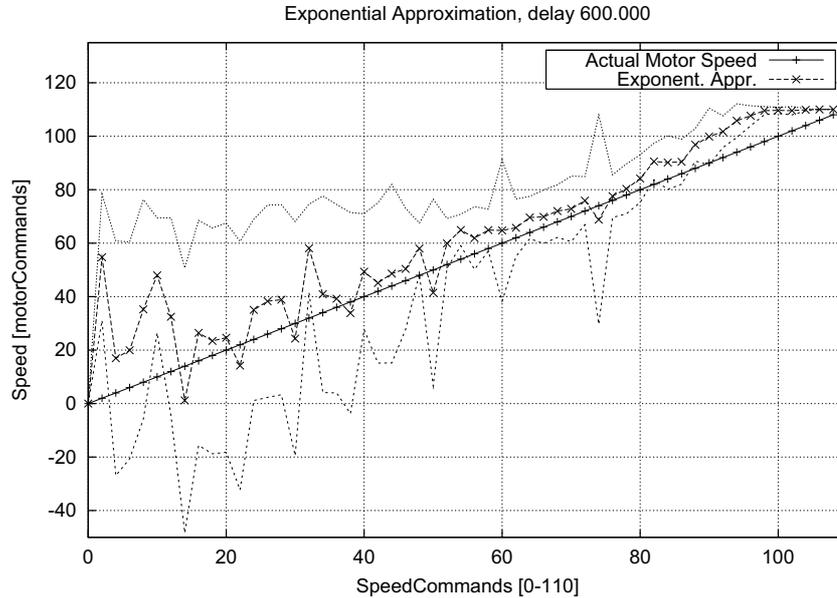


Figure 5.5: *Exponential function approximation Results.* The straight ('-+-') line is the actual motor command, and the exponential approximation curve ('-x-') is the mean of the estimated motor command produced by the exponential function approximation method. The two small dashed lines represent the standard deviation of the estimated motor command, obtained by repeating the experiment a hundred times.

In fig. 5.5 we present the results achieved with this method. In this figure the straight ('-+-') line is the actual motor command, and the exponential approximation curve ('-x-') is the estimated motor command. The small dashed lines represent the standard deviation of the estimated motor command, obtained by repeating the experiment a hundred times. As we can see in the graph, the estimation of the actual motor command is rather exact, but the standard deviation, especially in the lower-speed region, is quite high. This can have two causes. First, the motors may shake a little for slow-speed commands, this resulting in a shaky unreliable optic flow pattern. We did not pinpoint the exact cause of the problem. Second, the  $\alpha$ -value might be too high, so that the low-pass filter is not empty before the EEMD-module is presented to a new speed command. This results in disturbing noise.

### 5.2.3 The Artificial Neural-Network Curve Fitting

This approach uses a neural network to translate the EEMD-output to the appropriate speed command. The goal of this approach was first of all to achieve a good approximation and second, to minimize the computational load. We solved this by implementing two very small neural networks, each one learned a part of the EEMD-curve similar to the two parts that were approximated in the linear approach. The reason for using two small networks was that the bigger the network, the more calculation per time-step and therefore a drop in frame-rate which results in a diminished gaze stabilization performance. Eventually we used two neural networks of the form

presented in fig. 5.6, which we trained with the back-propagation algorithm (Haykin, 1999).

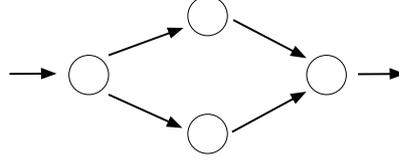


Figure 5.6: The neural network used in the function approximation, it represents the simplest network possible, containing one input neuron, two hidden neuron and a single output neuron.

### 5.2.4 Summary of the Back-Propagation Algorithm

The training method used for training the two small neural networks is according to the back-propagation rules.

#### Initialization

The weights and thresholds are initialized according to a uniform distribution whose mean is zero and whose variance is chosen in such a way that the standard deviation of the weighted sum of all synaptic inputs plus the bias lies in the transition between the linear and saturated parts of the sigmoid activation function

#### Presentation of the Training Examples

We present the network with an epoch of training examples. For each example in the set, ordered in a random order, the sequence of forward and backward computation, described in the following points were performed.

#### Forward Computation

A training example is denoted by  $[x(n), d(n)]$ , with  $x(n)$  as the EEMD-output applied to the input-layer and desired response  $d(n)$ , the stabilizing speed command. The activity  $v_j^{(l)}$  for neuron  $j$  in layer  $l$  is:

$$v_j^{(l)}(n) = \sum_{i=0}^{m_0} w_{ji}^{(l)}(n) y_i^{(l-1)}(n) \quad (5.7)$$

where  $y_i^{(l-1)}(n)$  is the output (function) signal of neuron  $i$  in the previous layer  $l - 1$  at iteration  $n$  and  $w_{ji}^{(l)}(n)$  is the synaptic weight of neuron  $j$  in layer  $l$  that is fed from neuron  $i$  in layer  $l - 1$ . For  $i = 0$ , we have  $y_0^{(l-1)}(n) = +1$  and  $w_{j0}^{(l)}(n) = b_j^{(l)}(n)$  is the bias applied to neuron  $j$  in layer  $l$ . The output signal of neuron  $j$  in layer  $l$  is:

$$y_j^{(l)} = \phi_j(v_j(n)) \quad (5.8)$$

where  $\phi(x)$  represents a sigmoid-function. If neuron  $j$  is in the first layer (i.e.,  $l = 1$ );

$$y_j^{(0)}(n) = x_j(n) \quad (5.9)$$

where  $x_j(n)$  is the  $j$ th element of the input vector  $x(n)$ . If neuron  $j$  is in the output-layer (i.e.,  $l = L$ , where  $L$  is the depth of the network)

$$y_j^{(l)} = o_j(n) \quad (5.10)$$

where  $o_j(n)$  represents the network output. The error signal is computed:

$$e_j(n) = d_j(n) - o_j(n) \quad (5.11)$$

where  $d_j(n)$  is the  $j$ th element of the desired response vector  $d(n)$ .

### Backward Computation

The computation of the local gradients of the network is defined by:

$$\delta_j^{(l)}(n) = \begin{cases} e_j^{(L)}(n)\phi_j'(v_j^{(L)}(n)) & \text{for neuron } j \text{ in out-layer } L \\ \phi_j'(v_j^{(l)}(n)) \sum_k \delta_k^{(l+1)}(n)w_{kj}^{(l+1)}(n) & \text{for neuron } j \text{ in hid-layer } l \end{cases} \quad (5.12)$$

where the prime in  $\phi_j'(\cdot)$  denotes differentiation with respect to the argument. Adjust the synaptic weights of the network in layer  $l$  according to the generalized delta rule:

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha[w_{ji}^{(l)}(n-1)] + \eta\delta_j^{(l)}(n)y_i^{(l-1)}(n) \quad (5.13)$$

where  $\eta$  is the learning-rate parameter and  $\alpha$  is the momentum constant.

### Iteration

The forward and backward computation are repeated by presenting new epochs of training examples to the network until the stopping criteria is met; the absolute rate of change in averaged square error per epoch is smaller than 1.0 %

$$\mathcal{E}_{av}^2 = \frac{1}{N} \sum_{n=1}^N (d_j - o_j)^2 \quad (5.14)$$

where  $\mathcal{E}_{av}^2$  represents the averaged square error energy.

### 5.2.5 Results of the Neural-Network Curve-Fitting Method

Figure 5.7 presents the results achieved with the curve fitting network based on a neural network method. The two small networks are trained according to the rules explained in the latter section. The method was tested in the way the exponential curve fitting method was tested. The motor was swung from left to right with different speeds, while our system estimated the given motor command. In fig. 5.7 the straight line ('-+-') is the actual motor command, and the '-x-'-line represents the mean output of the neural network curve fitting method. The experiment was repeated 100 times in order to get a good impression of the performance of the system. The small dashed line is the  $\pm$  standard deviation of this signal. As shown in the graph, the estimation of the given motor command is good. Another thing we can deduce from the graph is whether noise is present in the lower speed regions.

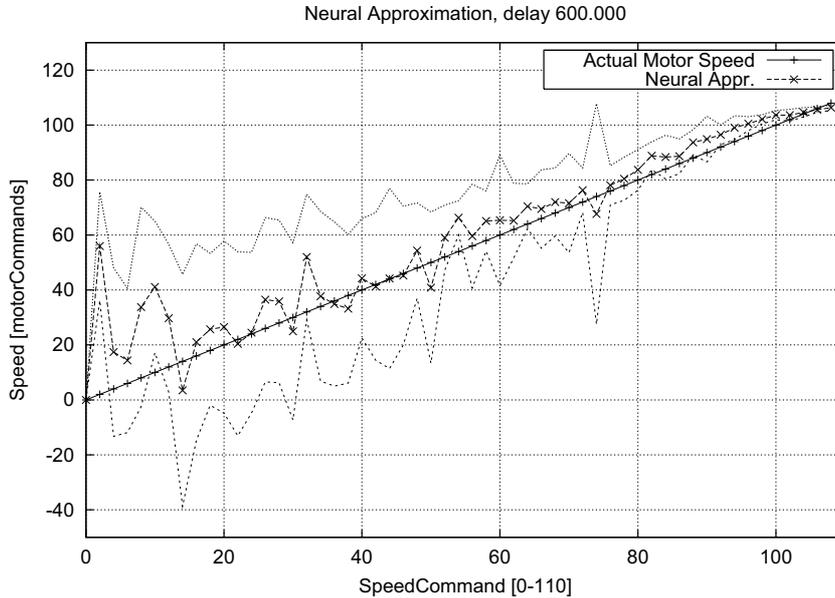


Figure 5.7: The Neural Network function approximation results. The straight ('-+-') line is the actual motor command, the '-x-' line is produced by the neural network function approximation approach. The two dashed lines represent the standard deviation ( $\pm\sigma$ ). As we see, this approach is able to estimate the given motor command, although the  $\sigma$  shows the presence of noise in the system. Note the noise in the lower speed region.

### 5.2.6 Comparing the Results

To compare the two curve fitting methods, the root mean square (RMS) is applied. The RMS-error has the same scaling as the curves have in fig. 5.7.

$$\mathcal{E}_{\text{rms}} = \sqrt{\mathcal{E}_{av}^2} \quad (5.15)$$

Comparing figure 5.7 with figure 5.5, we see many resemblances. But in table 5.1 the RMS error (equation 5.14) shows that the neural network approach is better than the exponential method, although this is only a minimal difference. The computational load for instance is also approximately equal. This data is obtained by calculating the RMS error of the estimated motor command, estimated by one of the two curve fitting methods, minus the actual given motor command (see graph 5.7 and 5.5).

Curve Fitting Method	$\mathcal{E}_{\text{rms}}$
Exponential	12.85
Neural network	11.13

Table 5.1: This table shows the average square error of the exponential and the neural network curve fitting. This data-comparison shows that the neural network method is slightly better than the exponential curve fitting method.

The reason we also implemented the neural network method is that in the future this

learning method is to be used for adjusting the network online. This is necessary to cope with sudden changes in the environment (indoor/outdoor light intensities) and it will result in a robust performance. Also, this method does not rely on manually determined parameters and is therefore easier to rehearse.

### 5.3 Gaze Stabilization Results

The neural network approach and the exponential approach show rather good approximation results as shown in the latter sections. In this section we present the gaze stabilization results based on the exponential and the neural-network approach in the robot dog. To test our system we moved the head of the robot from left to right. One camera of the robot head was fixed to its head, while the other was performing stabilizing movements in order to keep the image in the camera stationary. We tracked the same small object in both camera images, the absolute pixel-location of this object was plotted in figure 5.8.<sup>1</sup>

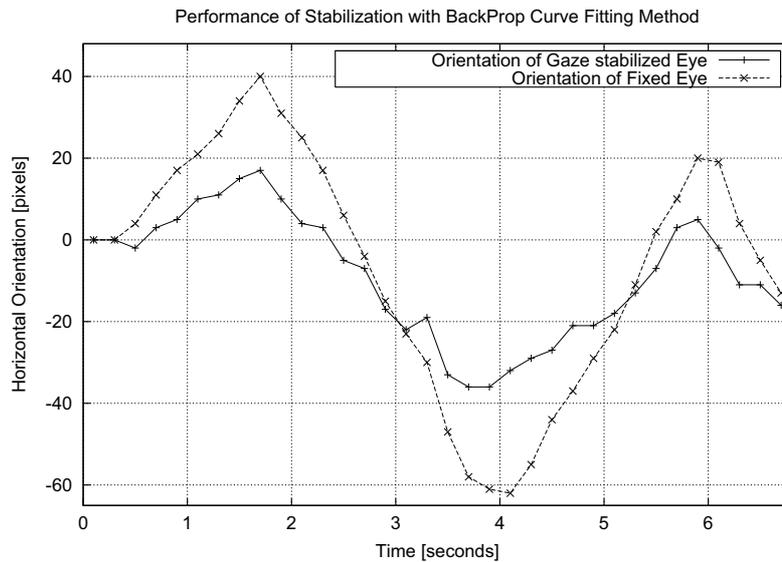


Figure 5.8: *Gaze stabilization results based on the back-propagation trained neural network, that learned what EEMD-output corresponded to motor command, in order to sent stabilizing commands to the motors. The '-+-' line represents the view-direction of the stabilizing eye. The '-x-' line is created by the eye fixed to the head, and therefore shows the trajectory of the head itself. The gaze stabilizing eye performs stabilizing movements (see text for details).*

We conducted a performance test of the proposed gaze stabilization method, in which a simple neck/head behavior was programmed by using 6 motors. The target behavior of neck/head system was prepared, i.e. the swing motions between right and left. The actuation of the neck/head motors was not sensory motor coordinated, but simply supplied constant energy (constant duty ratio), and changed the direction after a cer-

<sup>1</sup>Movies of the gaze stabilization behavior can be acquired by sending an email to [rick@ai.rug.nl](mailto:rick@ai.rug.nl)

tain period of time. One of the camera systems is simply fixed on the head, and is used to analyze the motion of the head in later stage, whereas the other camera system stabilizes the gaze by using the model described in the previous section (Fig. 4.2). After the experiments, the performance is evaluated by motion of the stored image, in which a fixed reference point in the environment was tracked. This information provides angular displacement of the image during the computational time steps, which corresponds to the ego-motion of cameras. Namely, while the neck/head system in action, if the trajectories of this reference point in the image stays at a point, gaze stabilization is successful. The experiments were conducted in an uncontrolled office environment. The sample snapshot from the robot perspective is shown in figure 5.9, captured by the active vision system. The frame-rate of the image processing is fixed to 20 frames per second.

Figure 5.8<sup>2</sup> illustrates two trajectories, one from the fixed camera and the other from the camera with gaze stabilization based on the back-propagation curve fitting method. The values in the vertical axis indicate the angular distance of the external reference point with respect to the camera image. Therefore the angular velocity of the ego-motion can be obtained by the corresponding derivatives in this figure.

## 5.4 Discussion

Due to the non-linear dynamics of the musculo-skeletal structure, the motion of the head shows relatively complex behavior, which make gaze stabilization very difficult. One explanation of the sub-optimal gaze stabilization results is that the elasticity of the spring factor in the muscles produces a totally non-linear force, irrespective of the neck-motors constant velocity. Another explanation is that because of the lack of accuracy of the lower speed region of the function approximation, this module does not always generates the proper velocity command. In this gaze stabilization process, this gives rise to an incorrect compensating movement. The implication of this fault is that visually-based gaze stabilization generates its “own” image flow, while it is compensating for another source of ego-motion. The system sometimes reacts to motion elicited by itself.



Figure 5.9: A snapshot of the uncontrolled office environment, which is used in the experiment. The image is captured by the active vision system.

In the gaze stabilization experiment higher optical flow velocities do occur. Although the extended model of motion detection enhances the range of the optical flow velocity compared to the single EMD, it does not result in perfect gaze stabilization behavior.

<sup>2</sup>Appendix C contains the gaze stabilization graph based on the exponential curve fitting method

An obvious solution to increase the performance of the robot dog appears to be the addition of another EMD in the extended elementary motion detector. This additional EMD would have a bigger distance between its photoreceptors, which would result in a detector able to measure higher optical flow values. But the motion detector is able to measure optical flow up to a maximum of 70 deg/s, which corresponds to a motor command of 120. As we see in appendix B, the real motor speed increases non-linear as the motor command increases. In the higher speed command region, we lack expression power to correct the high speed optical flow exactly, therefore we did not generate speed command higher than 110. This value was manually determined.

In biological systems as well as in the robot dog, it takes time to process the visual information which causes a latency to occur. This is the reason for the performance of the optokinetic reflex to decrease as the image speed increases. The system should react faster in this situation, but the visual processing speed (CPU-speed) is a bottle-neck. This conflict is also present in the robot dog: The frame-rate of the frame-grabber had to be limited to a fixed 20 Hz. Which means that very fast image motion (above approximately 70 deg/s) will appear as a blur and thus the robot is not able to cope with such fast motion. The solution to this problem would be to increase the frame-rate of the frame-grabber and the CPU-speed of the visual processor. The maximal perceivable image flow velocity does not depend on the detector but depends more on the frame-rate of the frame-grabber. With an increasing frame-rate, the EEMD can be modified to detect higher velocities.

An argument against this enhancement is how fast this optokinetic reflex actually should be. As mentioned earlier (section 2.6), since the vestibulo-ocular reflex in nature considerably contributes to stabilize the gaze during the high speed head movements, visually-based gaze stabilization might not need to cope with these high speed image velocities. In the natural habitat of the dog image flow up to 70 deg/s occurs for instance as the dog walks or nods its head. Higher image flow velocities occur when the dog is running and especially when the dog jumps. The additional enhancement this system needs is an implementation of the vestibulo-ocular reflex. This reflex is able to detect head velocities, much greater than our recent implementation of the optokinetic reflex. Therefore we can say that our implementation of the optokinetic reflex is successful. We have created a motion detector able to measure optical flow in a deterministic manner, and we used this detector to stabilize the gaze of the robot dog.

Compared to other existing artificial intelligence projects mentioned in chapter three this project differs from those projects on several aspects. First, this is not a humanoid project, but a canine project. Second, the robot dog project has its focus on locomotion, while the other artificial projects work with a stationary robot torso. These projects do not include topics like locomotion. Third, the gaze stabilization approach used in the robot dog project is characterized by biological plausibility and the goal is to build a fully plausible mimicked mechanism of the canine's stabilization mechanism, starting with the optokinetic reflex. The four existing projects have implemented an analogous version of the optokinetic reflex based on a non biologically inspired mathematical correlation method to calculate image motion.

## Chapter 6

# Conclusion

Visual perception of the surrounding world is an important source of information for the survivability of an animal. The problem in using this source is that motion, either elicited by the surroundings or by the animal itself, disturbs the source. The task of the visual system is to cope with wildly moving projections of the surroundings on the animals retina. When the disturbances are caused by ego-motion, this process is called *gaze stabilization*. Gaze stabilization has the goal of keeping the image, which is projected on the retina, stationary while the animal itself disturbs the projection by locomotion. Good examples of robust autonomous systems are found in nature. Their efficient way of problem solving can be used in the creation of robust artificial systems. At the AI-lab in Zürich the philosophy is to strive for a good comprehension of the internal mechanism that produces this robust behavior, a methodology characterized as *understanding by building*. This project was initially inspired by the efficient locomotion of the biological dog. The project's goal is to understand the perceptive processes which are the consequences of the physical structure of the dog. Therefore in designing the *robot dog*, we mimicked the dog's physical properties. The first task, achieving locomotion, is directly coupled to my graduation project: "Biologically inspired gaze stabilization" in which I examined the coupling of the eye motors to provide a stable view while the robot is walking.

### **What are the mechanisms behind gaze stabilization during ego-movements?**

In biological systems gaze stabilization is achieved by the cooperation of several reflexes:

1. The optokinetic reflex (OKR), which relies on visual information, i.e., translational and rotational optic flow, measured in the entire retina. The aim of the OKR is to keep the image still on the fovea. The OKR functions as a visually-driven closed-loop negative feedback tracking system.
2. The vestibulo-ocular reflex (VOR), which relies on signals produced by the vestibular organ in the middle ear. This organ determines the absolute movement of the head in three dimensional space, i.e., three linear- and three rotational movements.
3. Three supporting reflexes; vestibulo-collicular reflex, opto-collicular reflex and the cervico-ocular reflex.

Insects stabilize their gaze, while flying and turning, by moving their head faster than the thorax. Hereby is the time-span of the head-saccade minimized. Thus the stationary and stable view is maximized. In birds a similar approach can be observed, a walking pigeon bobs with its head making fast head-saccades to reduce the interruption

of stable gaze. This phenomenon is called saccade because of the resemblance between the jump-like eye-movements which are called saccades. Generating a conflict between the vestibular and the visual signals by positioning the bird on a stationary perch surrounded by an oscillating environment results in the behavior that the head moves along with the surroundings. This shows the visual predominance in the gaze stabilization process. Chameleons scan their environment with saccadic eye movements that are independent in the two eyes. Once the prey is located, the head is aligned and the two eyes converge to fixate the target binocularly. The chameleon's *gaze stabilization* behavior is characterized by fast eye-saccades, while its head, if needed, makes smooth pursuit movements. The interesting detail of the rabbit is that it lacks a fovea, which means that this animal provides the opportunity to study gaze stabilization in a strictly *isolated* form. In humans reflex-based gaze movements are often performed using eyes only. Humans have a frontal eye position, therefore information about distances of objects can accurately be obtained.

The goal of gaze stabilization is to minimize two types of optical flow on the retina. The first is rotational optical flow, which is elicited when the observer rotates along one of its three axes. The second is translational optical flow, which occurs when a passive observer undergoes a pure translational movement. Now that we know how a stable image is realized in biological systems, what does this information tell us in building artificial systems? Well, biological and artificial systems that share the same environment may adapt similar solutions to cope with similar problems.

### How can these mechanisms be realized in artificial systems?

Gaze stabilization has the goal to minimize retinal slip. The vestibulo-ocular reflex (VOR) can be implemented using a gyroscopic device, which simulates the functionalities of the vestibular organ. In this project we first focus on the optokinetic reflex (OKR), as a first step toward a fully-integrated gaze stabilization model. Other existing artificial system projects, also inspired by biological solutions, also integrated the two reflexes to build a robust gaze stabilization system.

To calculate the optical flow in the image, we used a biologically plausible model, that has been developed by Reichardt based on the directionally selective ganglion cells found in insects and several vertebrates. This model is named the 'elementary motion detector' (EMD). The EMD consists of two photoreceptors, low-pass filters, multiplication and a subtraction. The motion signal is calculated as follows. Two spatially adjacent photoreceptors receive luminance information. These signals are delayed by a temporal low-pass filter, a good approximation that biology also appears to use. The delayed channels are then correlated with adjacent non-delayed channels by means of a multiplication operation. Finally, the output of two anti-symmetrical half-detectors, each sensitive for motion in one direction, are subtracted to obtain a directional motion response.

In our implementation of the EMD we extended this model. Instead of using only two adjacent photoreceptors, we added pairs of photoreceptors with a wider distance. The maximum activation of the outputs of the extended elementary motion detector (EEMD) are selected as output. Compared to the original models, the EEMD provides a wider range of motion velocity measurement. The motion signal produced by the EEMD is used to stabilize the gaze of the robot dog. Thus gaze stabilization based on visual information, and is therefore analogous to the optokinetic reflex.

The design of the robot dog mimics the physical properties of the anatomy of the biological dog. In this project we mainly used the neck/head/eye-system. Six muscles (wires and springs) are installed in the frontal part of the dog, which control eight passive degrees of freedom in the spinal blocks. A binocular active vision system is placed in the skull of the robot. EEMDs cover the entire image along both horizontal and vertical axes. The outputs are spatially integrated, averaged and low-pass filtered. The corresponding motor command to generate the appropriate stabilizing behavior is calculated using two different types of curve fitting methods. First a mathematical approach, which uses the exponential form of the EEMD-curve to calculate the motor command and second an artificial neural-network approach, in which a back-propagation learning method adjusts the weights of two small neural networks. The average square error of the neural network method was lower, and thus this method is considered better. This method was used in successfully stabilizing the gaze of the robot dog. While the motion detector is able to measure optical flow up to 70 deg/s, the system cannot respond to optical flow of a higher order, because of image blur. This maximal image flow velocity can be increased by increasing the frame-rate of the frame-grabber. It seems reasonable to assume, that similar limitations exist in the biological gaze stabilization, necessitating an additional sensing signal apart from vision. In many species this signal is delivered by the vestibular organ. This organ measures head velocities higher than measurable by method based on optical flow. To obtain a perfect gaze stabilization performance the robot dog needs to be equipped with the optokinetic reflex and with a device with the functionality of the vestibular organ.

### **How does the proposed gaze stabilization help higher cognitive tasks such as object recognition?**

In this thesis we have discussed the need for a gaze stabilizing mechanism; The problem of relying on visual sensor information is that motion, caused by ego-motion, such as walking or running, disturbs this sensor. Consequently, a robot that walks or runs definitely needs to stabilize its gaze before it is able to do tasks such as object recognition. Although gaze can be stabilized, a visual perception system, will always have to deal with residual optical flow. This flow is in many cases not a nuisance, but a means to solve, e.g., 3D object-perception. Thus the extended elementary motion detector (EEMD), developed in this project, can also deliberately use image flow to detect objects.

A future topic might be object-tracking based on motion in a local area of image. Preliminary experiments with the robot dog show that just a small modification of the actual system, enables object-tracking with a stationary background. More sophisticated techniques with a non-stationary (counter-moving) background is definitely an interesting feature to explore.



# References

- Adelson, E.H. and J.R. Bergen. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am.*, 2:284–299.
- Amthor, F. and N.M. Grzywacz. (1993). Directional selectivity invertebrate retinal ganglion cells. In F.A. Miles and J. Wallman, editors, *Visual Motion and Its Role in the Stabilization of Gaze*, volume 5. Elsevier Science Publishers B.V., pages 79–100.
- Barlow, H.B. and W.R. Levick. (1965). The mechanism of directionally selective units in the rabbit's retina. *J. Physiol.*, 178:477–504.
- Borst, A. and M. Egelhaaf. (1989). Principles of visual motion detection. *Trends Neurosci*, 12:297–306.
- Borst, A. and M. Egelhaaf. (1993). Detecting visual motion: Theory and models. In F.A. Miles and J. Wallman, editors, *Visual Motion and Its Role in the Stabilization of Gaze*, volume 5. Elsevier Science Publishers B.V., pages 3–27.
- Breazeal, C. (1999). Imitation as social exchange between humans and robots. In *Proceedings of the 1999 Symposium on Imitation in Animals and Artifacts (AISB99)*, pages 96–104. Edinburg, Scotland.
- Bruce, Vicki, Patrick R. Green, and Mark A. Georgeson. (1996). *Visual Perception, Physiology, Psychology, and Ecology*. 3rd edition. 27 Church Road, Hove, East Sussex, BNA 2FA, UK: Psychology Press, Bookcraft Ltd., Bath, United Kingdom. ISBN 0-86377-451-2(Pbk).
- Burdess, C. (1996). *The vestibulo-ocular reflex: computation in the cerebellar flocculus*. Ph.D. thesis, University of Edinburgh.
- Collett, T.S., H-O. Nalbach, and H. Wagner. (1993). Visual stabilization in arthropods. In F.A. Miles and J. Wallman, editors, *Visual Motion and Its Role in the Stabilization of Gaze*, volume 5. Elsevier Science Publishers B.V., pages 239–264.
- Collewijn, H. (1977). Eye and head movements in freely moving rabbits. *Journal of Physiology*, (266):471–498.
- Draper, Mark. (1998). The adaptive effects of virtual interfaces: Vestibulo-ocular reflex and simulator sickness.
- Friedman, M.B. (1975). Visual control of head movements during avian locomotion. In *Nature*. London, pages 67–69.
- Fuchs, A.F. and M.J. Mustari. (1993). The optokinetic response in primates and its possible neuronal substrate. In F.A. Miles and J. Wallman, editors, *Visual Motion and Its Role in the Stabilization of Gaze*, volume 5. Elsevier Science Publishers B.V., pages 343–370.

- Gioanni, H., M. Bennis, and A. Sansonetti. (1993). Visual and vestibular reflexes that stabilize gaze in the chameleon. *Visual Neuroscience*, 10:947–956. Cambridge University Press.
- Goody, P.C. (1997). *Dog Anatomy: A Pictorial Approach to Canine Structure*. J.A., Allen & Co. Ltd.
- Gullan, P.J. and P.S. Cranston. (2000). *The insects, an Outline of Entomology*. Second edition. Chapman and Hall.
- Hateren, J.H. van. (1992). Theoretical predictions of spatio-temporal receptive fields of fly lmc8, and experimental validation. *J.Comp.Physiol*, (155):157–170.
- Haykin, S. (1999). *Neural Networks, a comprehensive foundation*. Second edition. Robbins, T.
- Hengstenberg, R. (1993). Multisensory control in insect oculomotor systems. In F.A. Miles and J. Wallman, editors, *Visual Motion and Its Role in the Stabilization of Gaze*, volume 5. Elsevier Science Publishers B.V., pages 285–298.
- Howard, I.F. (1993). The stability of the visual world. In F.A. Miles and J. Wallman, editors, *Visual Motion and Its Role in the Stabilization of Gaze*, volume 5. Elsevier Science Publishers B.V., pages 103–118.
- Ibbotson, Michael. (2001). identification of mechanisms underlying motion detection in mammals. In J.M. Zanker and J. Zeil, editors, *Motion vision, computational, neural, an ecological constraints*. Springer Verlag, Berlin, Heidelberg, New York.
- Iida, F. (2001). Goal-directed navigation of an autonomous flying robot using biologically inspired cheap vision. In *Proceedings of the 32nd ISR (International Symposium on Robotics)*, pages 1404–1409, 19-21 April.
- Iida, F. and D. Lambrinos. (2000). Navigation in an autonomous flying robot by using a biologically inspired visual odometer. In *Sensor Fusion and Decentralized Control in Robotic System III*, volume 4196, pages 86–97. Proceeding of SPIE, Photonics East.
- Kootstra, G.W. (2002). Selection of landmarks for visual navigation. Master’s thesis, University of Groningen.
- Kuniyoshi, Y., S. Rougeaux, O. Stasse, G. Cheng, and A. Nagakubo. (2000). A humanoid vision system from versatile interaction. In *BMCV 2000, LNCS 1811*, pages 512–526. S.W. Lee, H.H. Buelthoff, T. Poggio (Eds.).
- Lichtensteiger, Lukas and Peter Eggenberger. (1999). Evolving the morphology of a compound eye on a robot. In *proceedings of the third european workshop on advanced mobile robot*, pages 127–134. eurobot ’99, IEEE.
- McCollum, G. and R. Boyle. (2001). Conditional transitions in gaze dynamics role of vestibular nuclei in eye-only and eye/head gaze behaviors. *Biological Cybernetics*, 85:423–436.
- Metta, G. (2000). *Babybot - a study on sensori-motor development* -. Ph.D. thesis, LIRA-lab, DIST, University of Genova, Italy.
- Metta, G., R. Panerai, F. Manzotti, and G. Sandini. (2000). Babybot: an artificial developing robotic agent. SAB 2000, France.

- Oey, P.L. (1979). *Enige aspecten van het optokinetische onderzoek*. Ph.D. thesis, University of Rotterdam.
- Panerai, F., G. Metta, and G. Sandini. (2000). Vor-like stabilization reflexes in robots. In *8th European Symposium on Artificial Neural Network*, pages 26–28. (ESANN 2000), April. Bruges, Belgium.
- Pfeifer, R. and C. Scheier. (1999). *Understanding Intelligence*. Cambridge, Massachusetts.
- Pola, J. and H. Wyatt. (1993). The role of attention and cognitive processes. In F.A. Miles and J. Wallman, editors, *Visual Motion and Its Role in the Stabilization of Gaze*, volume 5. Elsevier Science Publishers B.V., pages 371–392.
- Pratt, D.W. (1982). Saccadic eye movements are coordinated with head movements in walking chickens. In *J. exp. Biol 97*. The MIT press, pages 217–223.
- Reichardt, W. (1969). Movement perception in insects. In W. Reichardt, editor, *Processing of Optical Data by Organisms and Machines*. Academic Press, New York, pages 465–493.
- Rogers, B. (1993). Motion parallax and other dynamic cues for depth in humans. In F.A. Miles and J. Wallman, editors, *Visual Motion and Its Role in the Stabilization of Gaze*, volume 5. Elsevier Science Publishers B.V., pages 119–138.
- Sandini, G., F. Panerai, and F.A. Miles. (2001). The role of inertial and visual mechanisms in the stabilization of gaze in natural and artificial systems. In J.M. Zanker and J. Zeil, editors, *Motion vision, computational, neural, an ecological constraints*. Springer Verlag, Berlin, Heidelberg, New York.
- Scassellati, Brian. (1998). A binocular, foveated active vision system. Technical report, Massachusetts Institute of Technology.
- Scassellati, Brian. (2001). *Foundations for a Theory of Mind for a Humanoid Robot*. Ph.D. thesis, Massachusetts Institute of Technology, May.
- Schilstra, C. (1999). Insect flight, eye movements and vision. Master’s thesis, BCN, June.
- Shibata, T. and S. Schaal. (2001). Biomimetic gaze stabilization based on feedback-error-learning with nonparametric regression networks. In *Neural Networks*. pages 201–216.
- Smakman, J.G.J., J.H. van Hateren, and D.G. Stavenga. (1984). Angular sensitivity of blowfly receptors: intracellular measurements and wave-optical predictions. *J. Comp. Physiol.*, (155):239–247.
- Srinivasan, M.V. (1993). How insects infer range from visual motion. In F.A. Miles and J. Wallman, editors, *Visual Motion and Its Role in the Stabilization of Gaze*, volume 5. Elsevier Science Publishers B.V., pages 139–156.
- Stancliff, Stephen B., Jennifer L. Laine, and Michael C. Nechyba. (1994). Learning to fly: Design and construction of an autonomous airplane.
- Tan, H.S. (1992). *Gaze stabilization in the rabbit*. Ph.D. thesis, Erasmus Universiteit Rotterdam.

Viollet, S and N. Franceschini, (1999). “Biorobotics, Methods & Applications”. chapter : Aerial minirobot that stabilizes and tracks with a bio-inspired visual scanning sensor. AAAI Press/ The MIT Press.

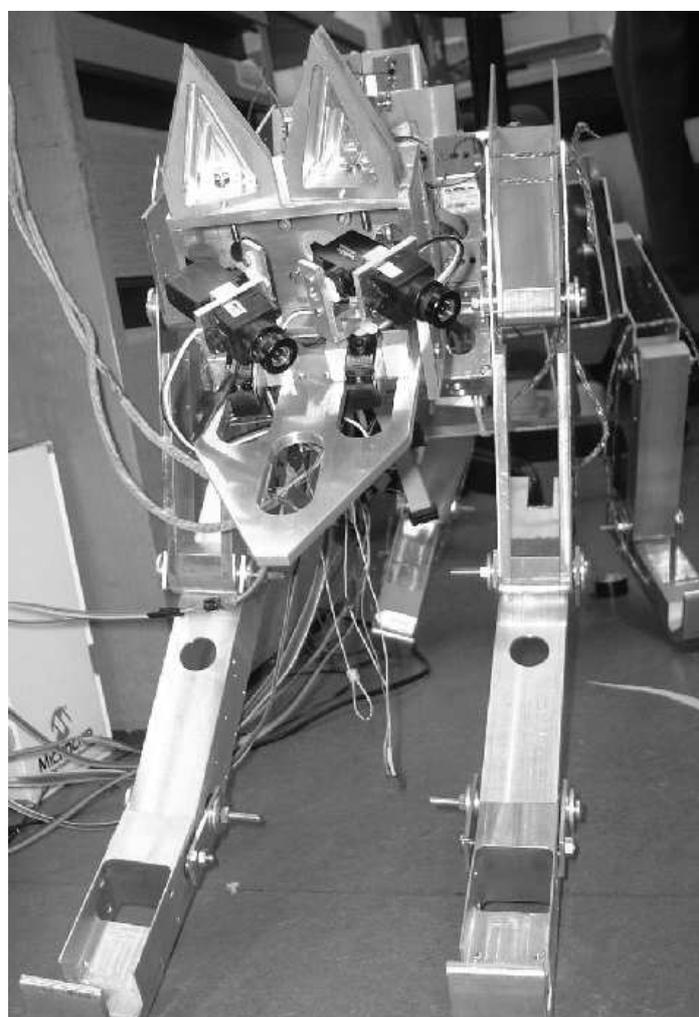
Wallman, J. (1993). Subcortical optokinetic mechanisms. In F.A. Miles and J. Wallman, editors, *Visual Motion and Its Role in the Stabilization of Gaze*, volume 5. Elsevier Science Publishers B.V., pages 321–342.

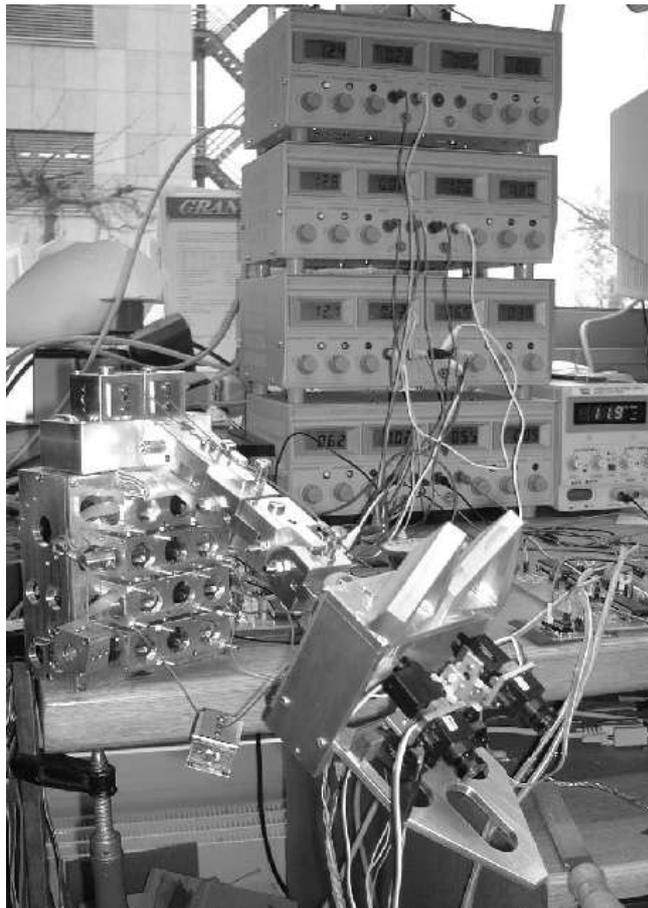
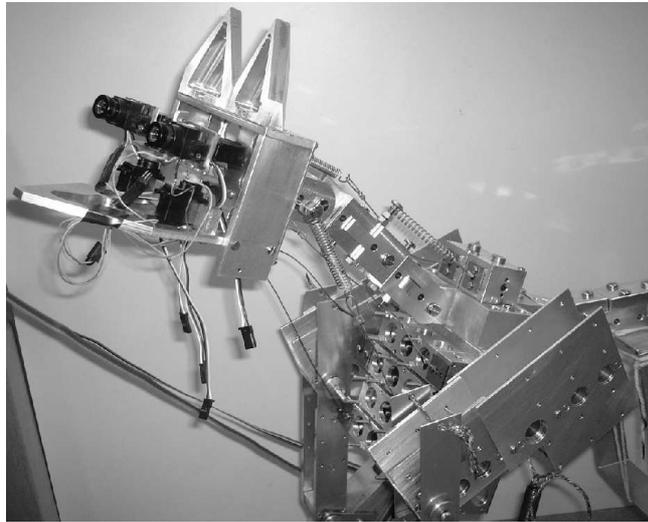
Wallman, Josh and Juan-Carlos Letelier. (1989). Eye movements, head movements, and gaze stability in birds. In *Vision, Brain, and Behavior in Birds*. The MIT Press, pages 245–263.

Walls, G.L. (1942). *The vertebrate eye and its adaptive radiation*. The Cranbook Press.

## Appendix A

### Pictures of the Robot Dog





## Appendix B

# Detection Range of the EEMD

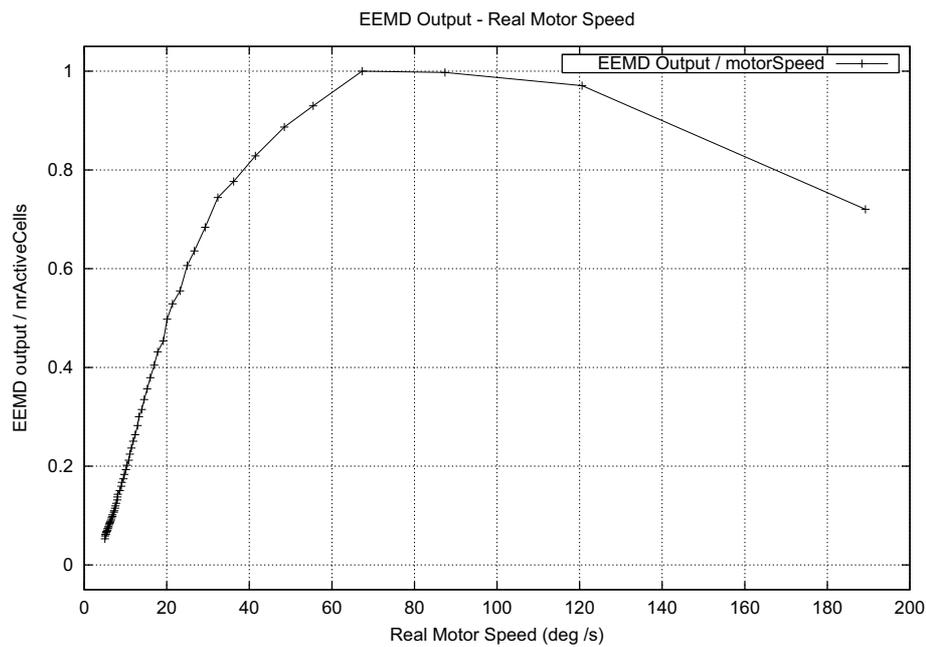


Figure B.1: *The range of the extended elementary motion detector (EEMD). We obtained this graph by plotting the graph of the characteristics of the motor (fig. B.2) against the EEMD-curve (fig. 5.3) showing the response of the EEMD to natural image motion. This graph shows the maximum image velocity the EEMD is able to measure, which is about 70 deg/s*

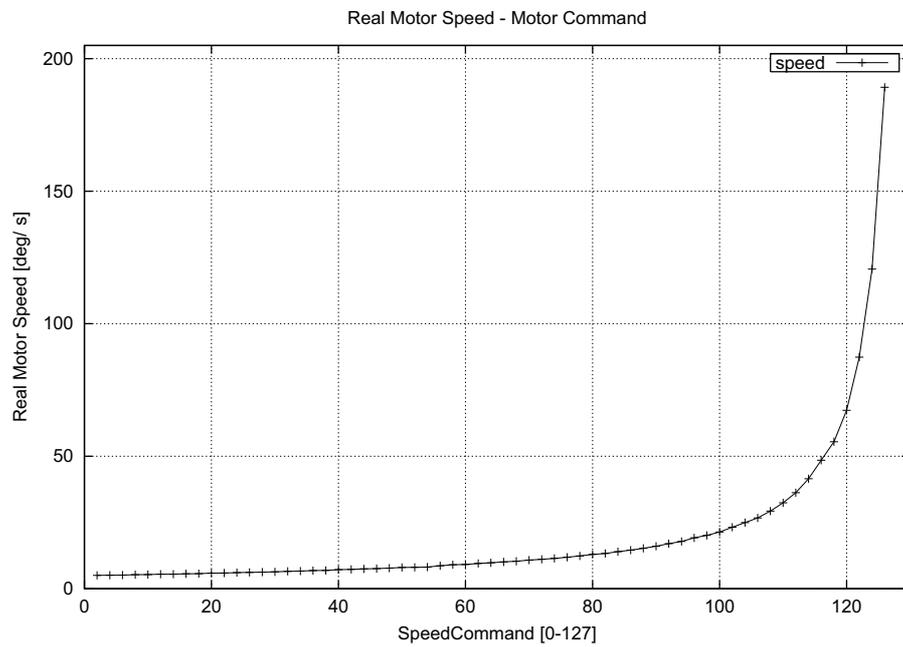


Figure B.2: *Characteristics of the motor. This graph was obtained by measuring the speed of the motor in degrees per second. It shows the non-linearity of the motor, a linear increase in terms of motor command, does not correspond to a linear increase of the actual motor speed. From this graph we can deduct the maximum image speed that the motion detection is able to measure. Using the maximum image velocity that the EEMD can measure from the latter graph (70 deg/s), we can deduct from this graph that this corresponds to the motor-command 120.*

## Appendix C

# Additional Gaze Stabilization Results

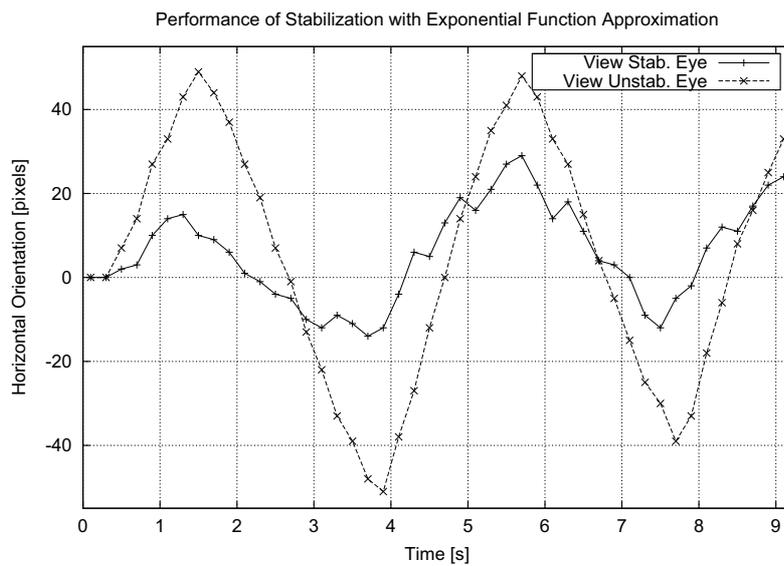


Figure C.1: *Gaze Stabilization Results based on the exponential curve-fitting method. The '-+-' line represents the view-direction of the stabilizing eye. The '-x-' line is created by the eye fixed to the head, and therefore shows the trajectory of the head itself. As we see, the gaze stabilizing eye performs stabilizing movements.*