

Visuo-inertial Stabilization in Space-variant Binocular Systems^{*}

Francesco Panerai^{a,1}, Giorgio Metta^b, Giulio Sandini^b

^a*Laboratoire de Physiologie de la Perception et de l'Action (LPPA), Collège de France, 11, pl. M. Berthelot, 75005 Paris, France*

^b*Laboratory for Integrated Advanced Robotics (LIRA), Department of Communication, Computers and Systems Science, University of Genoa, Via Opera Pia 13 - 16145 Genoa, Italy*

Abstract

Stabilization of gaze is a major functional prerequisite for robots exploring the environment. The main reason for a “steady-image” requirement, is to prevent the robot’s own motion to compromise its “visual functions”. In this paper we present an artificial system, the LIRA robot head, capable of controlling its cameras/eyes to stabilize gaze. The system features a stabilization mechanism relying on principles exploited by natural systems: an *inertial* sensory apparatus and images of *space-variant* resolution. The inertial device measures angular velocities and linear acceleration along the vertical and horizontal fronto-parallel axes. The space-variant image geometry facilitates real-time computation of optic flow and the extraction of first-order motion parameters. Experiments which describe the performance of the LIRA robot head are presented. The results show that the stabilization mechanism improves the reactivity of the system to changes occurring suddenly at new spotted locations.

Key words: inertial sensors, image stabilization, visuo-inertial integration, space-variant binocular vision.

^{*} The research described in this paper has been supported by the Italian Space Agency (ASI) and the EU Esprit project NARVAL.

¹ Corresponding author: Email: panerai@cdf-lppa.in2p3.fr, Ph: +33-01-44271389, Fax: +33-01-44271382

1 Introduction

Stabilization mechanisms have developed in natural systems as necessary functional elements to enable consistent treatment of visual information. These mechanisms are indeed important when the visual system needs to exploit its maximum acuity, estimate self-motion or discriminate motion of the visual scene. As a matter of fact, images falling on the retinas are required to be relatively still, particularly in all those tasks requiring the brain's neural circuitry to discriminate/extract/recognise relevant visual features. In humans, for example, it is well known that the stabilization of the whole visual field enhances sensitivity to individual moving objects and to motion parallax [27], contributing to enhancement in the perception of distance, form [35], and relative depth [12]. In robotics, too, image stabilization mechanisms should be considered important functional elements. Robots exploring the environment will certainly benefit from a stable gaze-line. Two cases are of particular interest: i) when performing dynamic visual measurements during their own movement [33]; ii) when using fixation as a reference while navigating or manipulating [3]. From an implementation perspective, how should one go about achieving robust performance and efficiency at the same time? What is the sensory information that should be considered to achieve the goal? Does the sensing process satisfy the dynamics of the stimuli? What are the computational cost of a particular implementation and the resources available to the robot ?

In animals with fixed eyes, like many insects and some birds, retinal image stabilization is achieved by compensatory head or body movements. Primates and many invertebrates with an efficient oculo-motor apparatus, rely mostly on compensatory eye movements. As a matter of fact, the “hardware” triggering compensatory motor responses is common to many biological species. A wide range of mechano-neural transducers, functionally equivalent to rotation or translation-sensitive mechanisms, are found in many species [48]. One can speculate about the advantages of these particular motion sensing “transducers”, but nevertheless it remains that such a particular design solution has been naturally selected to deal with the image stabilization problem. In primates, the mechanism controlling the direction of gaze on the basis of inertial information is called *Vestibulo-Ocular Reflex* (VOR). It is subdivided into *angular* VOR (AVOR) - generating oculo-motor responses to *angular* head motion - and *translational* VOR (TVOR) - generating responses to *linear* head motion [36,30]. In the case of the AVOR, the sensing is performed by three ring-shaped sensors (called semi-circular canals), measuring angular velocities along three perpendicular directions. In the case of the LVOR, the sensing is performed by the otoliths organs, which sense linear movements in horizontal and vertical directions and orientation of the head with respect to gravity [19]. The vestibular reflexes are known to operate in open-loop, are very rapid

and work best for high frequency movements of the head [20,5,48]. On the other hand, the visual reflexes, like the Opto-Kinetic Reflex (OKR), operate in closed-loop, are slower and respond better for lower frequencies of head movements [24,2].

The use of inertial information in artificial systems was already proposed in the past by several authors [45,4,31,7,23]. On the other hand, only few robotics implementations have appeared exploiting inertial cues for line-of-sight stabilization (e.g. [1]). The problem of image stabilization can be approached in different ways according to the structure of the system itself and to the nature of the environment. In recent work, Terzopoulos and Rabie prove that a “*pure visual*” approach is successful within their Animat world [44], a virtual world where autonomous virtual robots can perceive and interact in a physics-based environment. For the purpose of this paper though, the only interesting situation is characterized by real cameras moving in a dynamic environment. In this condition *stabilization* means being able to track a point in the environment and not simply being able to maintain a stable orientation of the camera’s optical axis. This goal can be achieved, in principle, through a process acting only on image features (see, for example, [11,28,38,47]). On the other hand, the goal of this work is to demonstrate that a more efficient solution is based on inertial measures providing angular and linear egomotion information independent of visual egomotion information derived, for example, through the optical flow field [15,16,22]. The inertial data, even if not sufficient to solve the stabilization problem, reduce the computation load of visual processing.

The paper is organized as follows: section 2 describes the inertial and visual sensory systems of the LIRA head; section 3 outlines some basic kinematics of gaze for translational and rotational movements. Section 4 proposes two control schemes developed in the framework of stabilization control: the first, with the intent to realize a monocular image stabilization, the second oriented to binocular systems. They both rely on the integrated use of inertial and visual information. In section 5 the experimental results and the stabilization performance of the LIRA robot are presented. The stimuli used to characterize the system’s response are externally imposed rotational and translational movements. Finally, the advantages of using a visuo-inertial stabilization mechanism in the context of fast coordinated eye-head movements are also addressed.

2 Inertial and space-variant sensing

The inertial sensory system for the robot head has been entirely designed and realized at LIRA Lab. This sensory module is designed according to the principles of the corresponding biological sensor, the vestibular system. The device

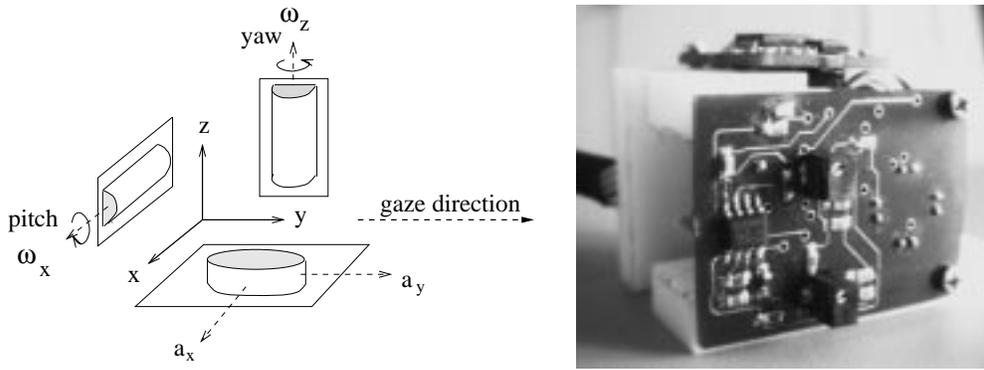


Fig. 1. The artificial vestibular system. **Left:** the spatial arrangement of the rotational and translational sensing elements. **Right:** the prototype artificial vestibular system. Overall dimensions are kept relatively small (5x4x4 cm) by implementing the design of the individual modules using surface mount technology.

is able to measure two angular velocities and two linear accelerations along two perpendicular directions. It is composed of modular, independent, sensing elements packaged together with the electronics for filtering and signal conditioning. The sensing modules are positioned as indicated in Fig. 1 (left); on the right side, the prototyped device is shown. The overall dimensions are 5x4x4 cm. The rotational part is composed of piezoelectric sensors (also known as vibrating gyros) produced by MURATA Inc. These transducers can measure rotational velocity around their longitudinal axis: the range of measurement extends from -90 deg/s to $+90 \text{ deg/s}$, the bandwidth from DC to 50 Hz (90 deg phase delay). The output of the sensor is band-pass filtered ($0.03 - 5.0 \text{ Hz}$) and amplified by custom electronics. The linear part, on the other hand, is composed of sensing elements exploiting a heat transfer design principle and are produced by PEWATRON Inc. The output generated is proportional to linear acceleration. The range of measurement extends from $-1g$ to $+1g$. In the linear part too, custom electronics is used to low-pass filter and amplify the response of the sensing elements [32].

As for the visual sensing, the LIRA's robot head acquires and processes images in a space-variant format (see [34] for a detailed description of the log-polar, space-variant geometry). Instead of a uniform resolution across the whole visual field, the robot eyes observe the environment through *foveal* image area of high resolution and a peripheral image area of low resolution. Fig. 2-left shows a space-variant, log-polar image as acquired by the head's vision system. The image (128x64 pixels) is magnified to better illustrate the resolution variability from foveal to peripheral areas. Fig. 2-right displays the corresponding Cartesian reconstructed image. These two biologically inspired sensory systems are integrated within a binocular architecture (see Fig. 3). More specifically, the active vision hardware used in these experiments is composed of:

- a four d.o.f. mount, independent vergence, common elevation and common pan,
- two color cameras simulating a space-variant (log-polar) geometry,



Fig. 2. The space-variant image geometry. **Left:** log-polar image of 128x64 pixels. **Right:** the corresponding reconstructed Cartesian image.



Fig. 3. The head mount. **Left:** front view of the robot head shows the offset position of the cameras with respect to the neck rotational axis (head geometrical parameters, $\mathbf{b}= 136$ mm, $\mathbf{a}= 95$ mm). **Right:** the artificial "vestibular" system is rigidly fixed on the back of the head mount.

- an artificial vestibular system,
- a double Pentium-Pro system architecture, a frame grabber (IntegralTech Flash Point 128) and a motion control board (Motion Engineering).

3 Head movement vs gaze kinematics

The goal of this section is to make explicit the relative roles of inertial and visual information in different situations. It should help to highlight the advantages of visuo-inertial integration and the asymmetries of eye-control commands. The description we develop here includes the eye-head geometric parameters and the eye-head kinematic dependence on fixation distance for rotations and translations. This formalism should not be taken to suggest that biological systems explicitly compute direct or inverse kinematics.

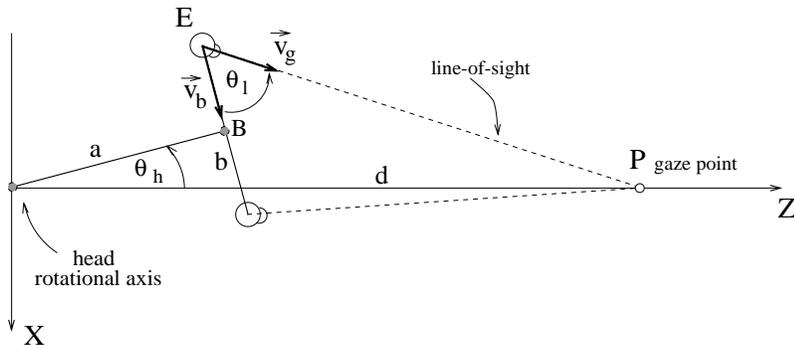


Fig. 4. Geometry of the eye-head system showing the parameters (\mathbf{a} , \mathbf{b} , \mathbf{d}) relevant to inertial and visual measures.

3.1 Rotational movements

We consider here the compensatory eye movements required to maintain stable fixation of a target at distance \mathbf{d} when the head rotates around a vertical, off-centered axis. Figure 4 shows the geometry of a binocular system for this case and indicates the most relevant geometrical parameters: the inter-ocular distance (or baseline) \mathbf{b} , the perpendicular distance \mathbf{a} between the rotational axis of the head and the baseline, and the viewing distance \mathbf{d} , measured from the head rotational axis and the gaze point. The analytical relation among these parameters can be derived by considering the kinematics of this model, and imposing the constraint that the eye \mathbf{E} maintains gaze at point \mathbf{P} when the head rotates (see the appendix A.1 for details). Simple vectorial rules and differentiation with respect to time leads to the following expression of angular velocity ω_l :

$$\omega_l = \left[1 + \frac{dZ_l - (a^2 + \frac{b^2}{4})}{d^2 - 2dZ_l + (a^2 + \frac{b^2}{4})} \right] \omega_h \quad (1)$$

where $Z_l = (a \cos \theta_h + \frac{b}{2} \sin \theta_h)$ represents the Z -coordinate of the left eye. Equation 1 determines the relationship between eye velocity, ω_l , and (i) the geometrical parameters of the eye-head system (i.e \mathbf{b} and \mathbf{a}) and (ii) the distance \mathbf{d} of the fixation point \mathbf{P} , for any given head velocity ω_h . In a robot vision system different choices of the \mathbf{a} and \mathbf{b} parameters (i.e. geometric configurations and dimensions) determine different shapes of the eye-head velocity relationship. Equation 1 also makes explicit the inverse dependence upon distance [33]. Interestingly enough, dependence upon distance is clearly evident in primate's RVOR responses [6,40,46,17,13,43]. These findings suggest that fixation distance might play an important role in simplifying/synthesizing efficient oculo-motor responses to rotational movements especially in the close range domain. In this regard, note that the eye velocity ω_l required to maintain fixation on near objects can be as much as twice the value of ω_h , for

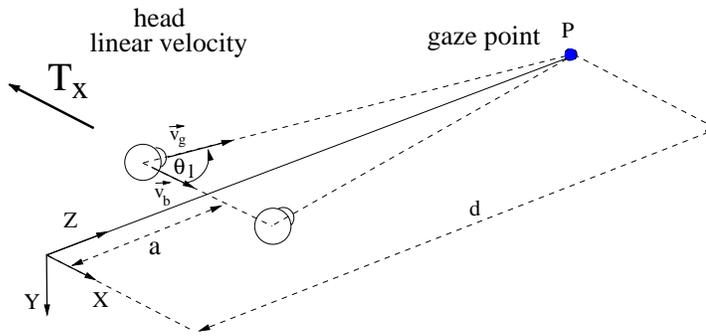


Fig. 5. Geometry of ocular responses to lateral translations.

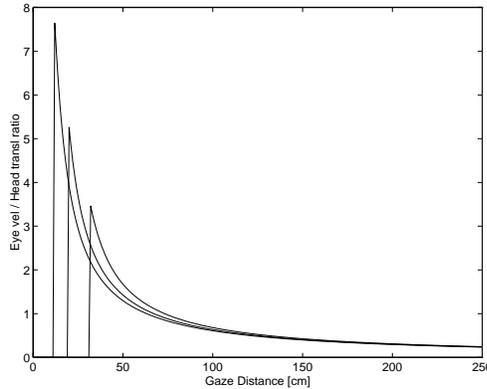


Fig. 6. Theoretical eye-head velocity ratios plotted for translations of the head. The velocity ratio as a function of fixation distance d for different values of a . The baseline b is fixed (6 cm) and eye-to-neck distance a increases (from 6 to 16 cm).

fixation distances in the range 25-200 cm; in other words, the optimal amount of ocular compensation needed to achieve a stable gaze can change rapidly with distance of fixation. One point worth stressing here is the fact that the range over which fixation distance influences the optimal control of the eye compensatory gain (at least from the kinematics point of view) may not be very relevant for locomotion (e.g. a robot walking/navigating and fixating at a long range). On the other hand this range overlaps entirely with manipulation workspace and, in this respect, might justify appropriate control circuits (e.g. a robot manipulating objects).

3.2 Translational movements

To understand the requirements of stable fixation in the case of translational movements, we refer to the same formalism used for rotations. Figure 5 shows the geometric representation of this task for a binocular system which is translating with instantaneous velocity T_x along the x -axis. The vectorial formalism

applied in this case gives :

$$\omega_l = \left[\frac{(d - a)}{\left(\frac{b}{2} - x\right)^2 + (d - a)^2} \right] T_x \quad (2)$$

Figure 6 (right) represents graphically the expression of the gain required for perfect compensation (i.e. equation 2). The curve shows clearly that the ocular response required to maintain fixation on near objects (i.e., 20-150 cm) can be quite large, but decreases inversely with distance. An object at infinite distance does not require, in principle, any ocular compensation irrespective of how fast translation is occurring. Moreover, when the eye is fixating an object in the range 50-200 cm, the required gain (i.e. required eye velocity per unit linear translation) changes with distance dramatically: from a value of 0.5 at a distance of 150 cm, the gain rises to about 2.0 at 50 cm. In terms of ocular velocity, this means that changing fixation from a point at 150 cm distance to another at 50 cm, while translating at velocity T_x , requires a four-fold increase in the eye velocity if fixation is to be maintained. Once again, it is interesting to remark that the gain of the primate TVOR and RVOR varies inversely with viewing distance, although compensation is often less than complete. The modulation of the TVOR gain is subject to instantaneous changes in fixation distance and it can be modeled using a linear relationship with binocular vergence angle, though vergence is not the only parameter used for range-finding [37,39].

3.3 Eccentric gaze

When looking at an eccentric target in near space, the compensatory eye movements required to maintain binocular alignment during head rotations are different for the two eyes [17]. From a kinematics point of view, the origin of this asymmetry is clear if one compares the analytical expressions of the two angular velocities (i.e. left vs right eye). The optimal compensatory velocities are given by equation 1 for the left eye and by the following for the right eye,

$$\omega_r = \left[1 + \frac{dZ_r - \left(a^2 + \frac{b^2}{4}\right)}{d^2 - 2dZ_r + \left(a^2 + \frac{b^2}{4}\right)} \right] \omega_h \quad (3)$$

where $Z_r = \left(a \cos \theta_h - \frac{b}{2} \sin \theta_h\right)$ represents the Z-coordinate of it. The expressions are almost identical except for the sign of a few terms encoding the opposite position of the eyes. The representation of the two gains in polar coordinates (see Fig. 7 - gains are plotted with respect to head angular position, θ_h , and for a given distance of fixation) reflects the asymmetric requirement of the optimal response. In the case of an object at 30 cm distance (Fig. 7-left),

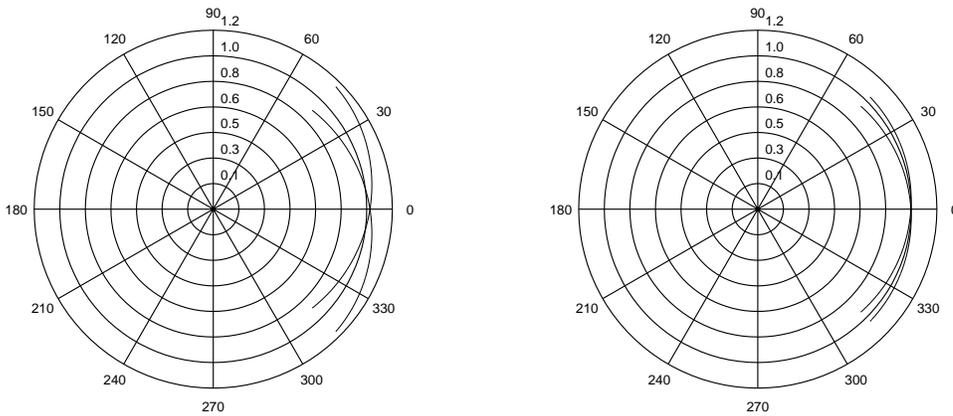


Fig. 7. Polar representation of the inertial gains (G_{vor_l} and G_{vor_r}) for the left and the right eye. The two curves in each plot describe the binocular gains required to perfectly stabilize the left and right image during a rotational movement. The angular displacement from the zero reference codes the orientation of the head while fixating frontally at a given distance. **Left:** distance of fixation is 30 cm. **Right:** distance of fixation is increased to 70 cm.

an angular deviation from the frontal direction of 30 deg introduces a relative gain difference between the two eyes of about 0.2; for example, with a head velocity of 200 deg/s , deviating 30 deg from the frontal direction gives a relative angular differential velocity of 40 deg/s . Thus, the angular velocities of the two eyes can be rather different in the near space. Although in humans there is clear evidence that during compensatory eye movements binocular alignment is not strictly maintained [10], this constraint might be more important for a robot vision system, especially if the system uses binocularly-derived cues to control camera movements [9].

4 Visuo-inertial integration

The issue of synthesizing simple and effective control strategies that integrate visual and inertial information appears challenging. In an anthropomorphic binocular system, the rotational axis of the eye and of the head do not coincide. Therefore, a rotational movement of the head causes both a rotation and a translation of the eyes. If the fixation point is at infinity, the inertial information alone will in principle be sufficient to stabilize gaze perfectly. However, this situation might represent a minority of cases. In general, as we have outlined in section 3, to generate an adequate compensatory response, the information about viewing distance is required. We suggest later in this section the use of a control strategy exploiting range information to tune the stabilization performance. Moreover, compensatory eye movements for translations also depend on other contextual factors, such as the direction of gaze with respect to the direction of heading [18].

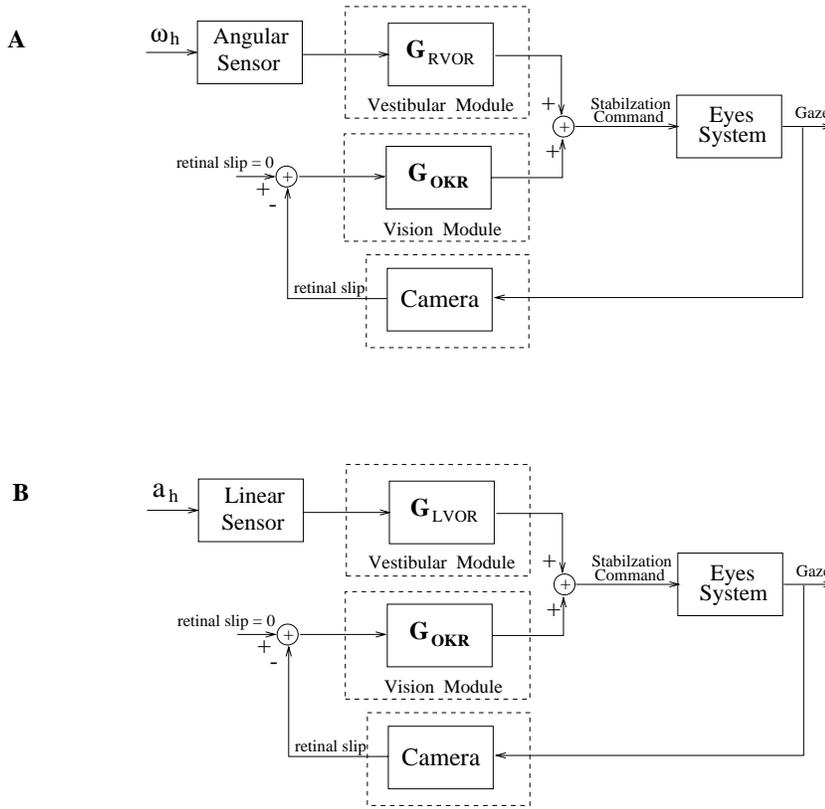


Fig. 8. Block diagrams of visuo-inertial mechanisms to stabilize robot's gaze. **A**: the rotational module selectively measures rotational movements of the head (ω_h). **B**: the linear accelerometer senses translational movements of the head (a_h). In both cases, the inertial information is processed open-loop and acts in parallel way with the vision module.

The structure of the basic visuo-inertial stabilization mechanism implemented on the LIRA head is sketched in Fig. 8. In the block diagrams, the visual and the inertial information are simply added together (in the biological literature this early modeling was named the "linear summation hypothesis"). The compensatory movement generated by the inertial information limits the amplitude of image motion to a range of values measurable by visual algorithms. Optic flow is therefore used to measure the residual error (called residual optic flow - ROF) present in each image after inertial compensation. In particular, with reference to Fig. 8 **A**, the stabilization command generated is:

$$SC(i+1) = G_{RVOR} \cdot \omega_h(i) + G_{OKR} * ROF(i)$$

whereas in the linear case (Fig. 8 **B**) the stabilization command is given by:

$$SC(i+1) = G_{LVOR} \cdot v_h(i) + G_{OKR} * ROF(i),$$

where $v_h(i)$ term represents an estimate of translational velocity as obtained from numeric integration of a finite number of $a_h(i)$ samples. Even if simply re-

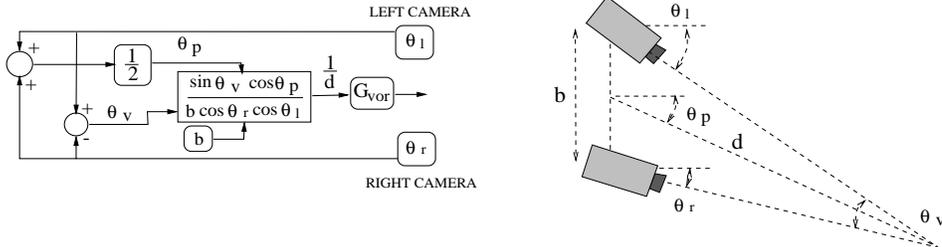


Fig. 9. Block diagram describing the computational scheme used to derive distance information.

alized, this stabilization scheme, and the experimental result obtained from it, draw the attention to the following important issues. First, the computational resources in an autonomous system are not unlimited. Second, the visual processing is computationally demanding: the amount of image data is usually much larger than many other sensory data; further, vision algorithms require, on the average, several pre-processing steps. In addition to that, whatever the algorithm adopted for image slip estimation, a limited range of image velocities can be reliably measured (we have estimated this limit for our system and results are summarized in appendix A.2). As a consequence, in a "real-world" environment, an approach to stabilization willing to consider visual information alone will have a relatively limited performance. In contrast, the control scheme described here enables the robot head, for example, to maintain a stable gaze at 1 m distance in response to transient rotations of the head at a speed of 10 *deg/s*, necessitating the eye/camera to rotate at 12 *deg/s* for full compensation; the inertial mechanism ($G_{vor}=1$) accounting for 10 *deg/s* of this compensation, leaving to the visual compensation mechanism to deal with a mere 2 *deg/s* residual flow. The advantage obtained is therefore clear: the use of inertial information enables more efficient use of the computational resources and it extends the range of motions or external disturbances the system can effectively deal with, without increasing system complexity. In biological systems the same sort of trade-off is present: vestibular reflexes have high-pass dynamics and are efficiently complemented by visual reflexes with low-pass dynamics [29,42]. Just to give a numerical example, experimental data on primate vestibulo-ocular responses indicate that they are ultra-rapid (RVOR latency is <10ms [41], and TVOR latency is <20ms [8]), while the complementary visual responses are considerably slower (>50 ms in monkeys [25] and >80 ms in humans [14]).

In order to work optimally also in a context where the fixation point is changed in depth, a tuning scheme which adapts the compensatory gains to changes in distance is proposed. The scheme exploits range information which is derived on the basis of dynamic control of vergence [9]. This type of control enables the system to keep both eyes/cameras pointing at the object of interest and to dynamically change their orientation in order to maintain the fixated object centered in the foveas. With reference to Fig. 9, the distance d of the object

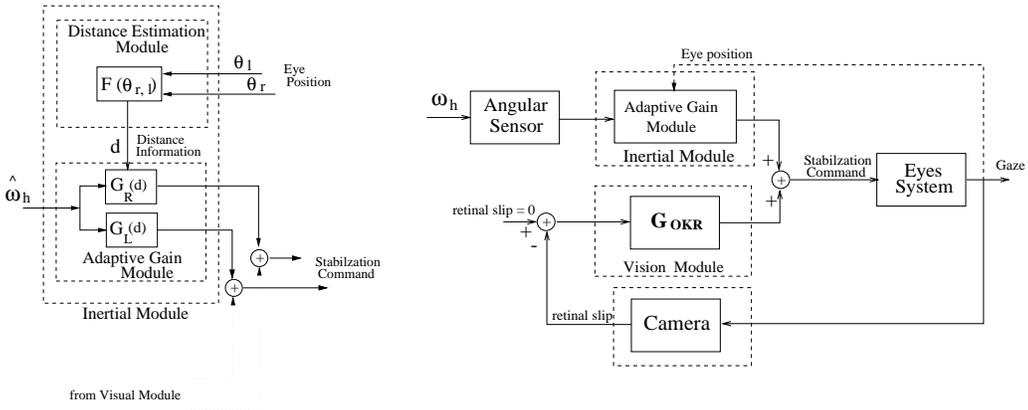


Fig. 10. Adaptive tuning of compensatory gains. **Left:** the compensatory gain of the left eye, G_L , and the corresponding gain for the right eye, G_R , are tuned according to distance of fixation. The block containing the function $\mathbf{F}()$ combines eye angular positions to derive an estimate of the fixation distance \mathbf{d} according to expression 4. **Right:** the adaptive gain module replaces the constant gain module of the previous scheme.

being fixated can be derived from the vergence (θ_v) and version (θ_p) angles, using the following equation:

$$d = b \frac{\cos(\theta_r) \cos(\theta_l)}{\sin(\theta_v) \cos(\theta_p)} \quad (4)$$

where $\theta_p = \frac{1}{2}(\theta_l + \theta_r)$ and $\theta_v = (\theta_l - \theta_r)$. Distance information is in turn used to adaptively change the compensatory gains of both eyes. The diagram in Fig. 10 describes the information flow which implements this scheme. The tuning functions we propose to use at this stage are those obtained from the kinematic analysis of subsection 3.3. A different approach would be that of learning these tuning functions.

5 Experiments

Four different experiments are presented. The first and second ones describe the stabilization performance in the monocular case: rotational movements around a vertical axis and linear translations along a fronto-parallel plane are addressed. The translational case will be described more in details. In fact, the stabilization performance for the rotational case has been extensively addressed in a previous work [33]. For both types of movements, the results show that slip on the image plane can be reduced and almost canceled by appropriate, constant tuning of the inertial (G_{vor}) and visual (G_{vis}) gains. The third experiment presents an adaptive tuning scheme for the eye

compensatory response for the translational case. The robot system estimates the distance of the fixation point and uses this parameter to tune the gains of the counter-rotational response. The last experiment shows to what extent the use of inertial information can improve image stability during coordinated eye-head movements. In all the experiments, stabilization performance is evaluated quantitatively by using one (in some case two) of the following visual measurements: correlation measurement between consecutive frames (also called image stabilization index - ISI), retinal error and residual optic flow (ROF). A detailed description of the technique used to estimate the ROF (from log-polar images) is given elsewhere [9]. On the other hand, the ISI index is defined as follows: $ISI_i = 1 - NC_i$,

$$NC_i = \frac{\sum_{\eta, \xi} (I_i(\eta, \xi) - \mu_i) \cdot (I_{i-1}(\eta, \xi) - \mu_{i-1})}{\sqrt{\sum_{\eta, \xi} (I_i(\eta, \xi) - \mu_i)^2 \cdot \sum_{\eta, \xi} (I_{i-1}(\eta, \xi) - \mu_{i-1})^2}} \quad (5)$$

The correlation measures the degree of similarity between two subsequent log-polar images (I_i and I_{i-1}). The symbols μ_i and μ_{i-1} indicate the corresponding image mean values. Better stabilization performance is mapped to lower values of ISI . During each experiment the following data are stored to enable an off-line evaluation of the overall performance: i) position angle of the inertial stimulus; ii) output of the inertial sensors; iii) the translational component of the optic flow; iv) values of the image stabilization index (ISI); v) position angle of the controlled camera; vi) retinal target position with respect to the center of the image (this index is used in particular in the binocular stabilization experiment). Position and inertial information are sampled at rate of 33 Hz (i.e. 30 ms interval). All the visual processing is performed at a sampling interval of 80 ms (i.e. 12.5 Hz). As regards the visual parameters measuring stabilization performance, the values reported in all the graphic presentations are obtained as statistics over several repetitive trials; for this reason data are often indicated along with the corresponding standard deviation.

5.1 Monocular stabilization: head rotations

The results of monocular image stabilization during head rotation around the vertical axis are here only summarized [33]. In this experiment the rotational axis of the inertial stimulus coincides with the common pan axis of the camera mount. The stabilization performance is evaluated in terms of the ROF measured at the center of the image plane. The parameters of the rotational stimulus (amplitude and frequency) were chosen to induce on the image plane a motion field close to the saturation threshold of the visual processing system. In our case this threshold is of the order of 4-5 pixels. Above this value, the relation between the image velocity and the optic flow is not linear (see the

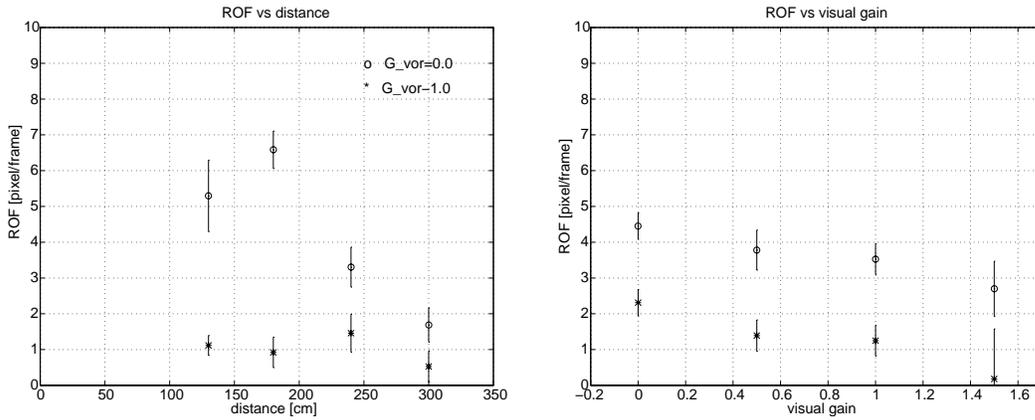


Fig. 11. Residual optic flow (ROF) on the image plane. **Left:** ROF for different distances when two different G_{vor} gains are used. **Right:** ROF for different values of visual gain G_{vis} (G_{vor} is constant: the 'o' symbol indicates measurements corresponding to $G_{vor} = 0.7$, while the '*' symbol indicates $G_{vor} = 1$).

appendix A.2). It is worth stressing that, even if the dynamics of the stimulus is not a “high dynamics” (i.e. 10 deg/s amplitude, 0.2 Hz frequency), in this case the pure visual approach could not be performed correctly (i.e. the optic flow processing is out of range). This experiment is performed in real-time with a control period of 80 ms². Two sets of ROF measurements are presented. The first, plotted in Fig. 11 (left) shows the decrease of the ROF with and without inertial stabilization as function of fixation distance. Note that in the case $G_{vor} = 1$, the ROF measured is always maintained within the range of correct measurements. It is in this case that ROF can be used to synthesize an additional motor signal to cancel out any remaining image slip. On the other hand, when the inertially generated response is purposefully removed ($G_{vor} = 0$), the values of the ROF obtained for a distance of 130 cm start to indicate saturation effects (i.e. the ROF drops to a lower value and the corresponding standard deviation indicates a larger variability of the measurement) generating unreliable information for the feedback loop. The second set of measurements (Fig. 11-right) show that when visual and inertial information are integrated, the performance of stabilization further improves. The two series of ROF measurements show a constant trend toward lower values when the the visual gain G_{vis} is introduced.

5.2 Monocular stabilization: head translation

This experiment describes the image stabilization performance of the LIRA robot in response to horizontal movements in the fronto-parallel plane. These

² The architecture used in this experiment did not support multi-tasking; this issue constrained the sampling interval of visual and inertial information to a common value.

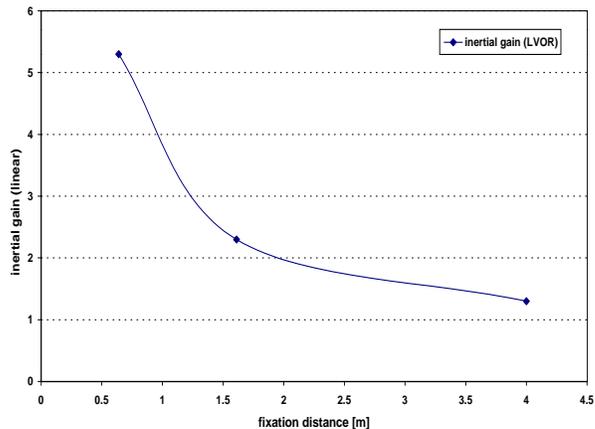


Fig. 12. Linear inertial gain as function of distance. Gain values are determined experimentally by minimizing the ISI visual measurement.

types of movements elicit the response of the linear part of the artificial "vestibular" system. Compensatory eye movements for fixation in the fronto-parallel perpendicular plane are addressed. The robot head is mounted on a support which is fixed to a computer controlled, translating table. This experimental arrangement enables to generate repetitive left-and-right movements with a desired acceleration profile. For the purpose of the current measurements, the cameras of the head are initially oriented to fixate a background plane; distance to the background, which is an experimental parameter, is changed throughout trials. The motion profile generated by the motorized table is trapezoidal, with an acceleration of 90 cm/s^2 and a maximum velocity of 30 cm/s . Three different fixation distances are considered: 64, 161 and 400 cm. For each distance an "optimal" tuning of the "linear" gain was obtained experimentally by minimizing the ISI index. The values determined for the gains are in the corresponding order, $G_{vor}=5.3$, 2.3 and 1.3. In Fig. 12 we have plotted such gain values as a function of the distance parameter (that is, distance to the background plane). Interestingly enough, the curve interpolating the experimental gains shows the same inverse dependence on distance as the "kinematic" curve derived in section 3.2. For each experimental trial, performance in stabilization is evaluated by calculating the residual optic flow (ROF) corresponding to the maximum translational velocity. Fig. 13 shows three ROF curves corresponding to the three different distances mentioned above. The left-most value in each curve represents the optic flow measured when no compensatory response is generated. Clearly enough, the higher the fixation distance (bottom value, $\mathbf{d}=4.00$), the lower the amplitude of the measured flow. When compensatory eye movements are generated (with the "optimal" *linear* gains), the ROF becomes smaller. Starting from each second value, the samples represent the ROF measured using the "optimal" inertial

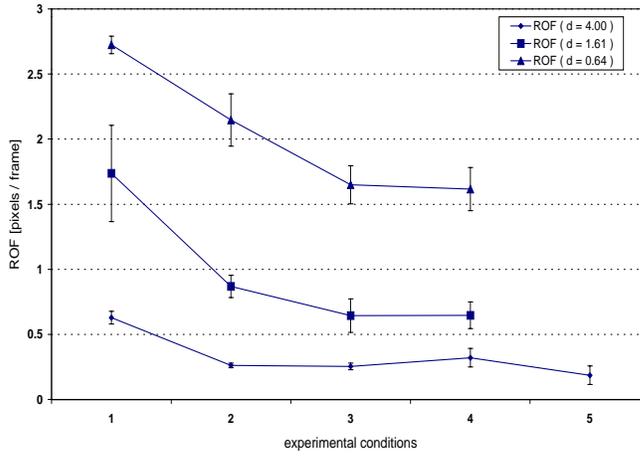


Fig. 13. Residual optic flow measured during head translations. Each curve refers to a different fixation distance (in cm) and a different G_{vor} constant value. **Top curve:** $d=64$, $G_{vor}=5.3$. **Middle curve:** $d=161$, $G_{vor}=2.3$. **Bottom curve:** $d=400$, $G_{vor}=1.3$. The experimental conditions correspond to a constant inertial gain and a different value of the visual gain for each fixation distance curve (from $d=400$ to 64 respectively). Conditions are: (1) $G_{vor}=0.0$, $G_{vis}=0.0$ for all the curves, (2) constant G_{vor} and $G_{vis}=0.0$, (3) constant G_{vor} and $G_{vis}=0.01, 0.07, 0.05$, (4) constant G_{vor} and $G_{vis}=0.085, 0.1, 0.07$.

gain G_{vor} and an increasing value of the visual gain G_{vis} (range of the visual gain used extends from 0.01 to 0.10). Although in two cases, namely $d=400$ cm and $d=161$ cm, stabilization performance improves with the appropriate tuning of the visuo-inertial response, the top curve ($d=64$ cm) shows clearly that for small distances image motion still remain high (in the order of 1.5 pixel/frame). At this stage, the limit for such small distances is, in our opinion, due to several factors: uncontrollable delays in the sensori-motor loop, phase lags introduced by sensor filtering (to eliminate the sensor noise), but most importantly, the low sensitivity of the linear sensing device. Further work is on-going to improve the linear part of the inertial sensory system.

5.3 Binocular stabilization: adaptive gain control

These experiments show how the robot system responds by adapting the oculomotor compensation strategy according to distance requirements. We have shown in section 3.1, that for rotational movements of the head a stable fixation can be obtained in the short and long range by tuning the compensatory gain inversely with distance. In section 4, we have proposed an adaptive con-

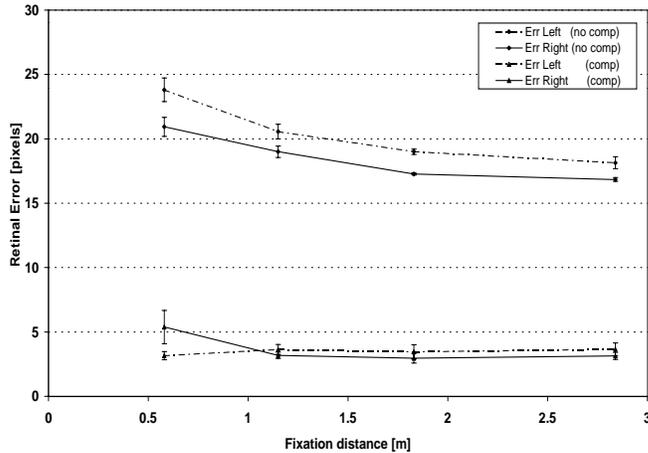


Fig. 14. Retinal error measured during head rotational movements. Upper curves measure error in the "non-compensated" case; lower curves represent error in the "adaptive compensation" case.

trol scheme exploiting distance information. Here we test the performance of this control strategy for several fixation distances. To obtain reliable distance information we have programmed the robot head to control its vergence (i.e. the distance of fixation point) on the basis of color information. An algorithm performing color-segmentation runs independently on each robot's eye. Position information of a target are extracted and used to keep the object centered in both images. A more general approach to control vergence would be the scheme proposed by Capurro et al. [9], which rely on binocular fusion and does not require a-priori knowledge of target characteristics. On the other hand, for the purpose of our experiment, the vergence control above mentioned is sufficient (and less computationally expensive) to derive consistent angular information relative to fixation point. When the target is foveated in each eye, the depth is calculated with equation 4. At the next stage, using the pure kinematic relations derived in section 3.1, the gains of the inertially-driven response can be adapted "on the fly" to the necessary requirements. A stimulus of constant amplitude and frequency (17 deg/s and 0.3 Hz) is used to impose sinusoidal rotational movements to the head. Stabilization performance is evaluated by means of the retinal error (in pixels) and of the ROF, at a sampling interval of 80 ms. Figures 14 and 15 show respectively the retinal error and the ROF for the "non-compensated" and the "compensated" case. The top curves of Fig. 14 describe the error without compensatory response. Note that when fixation distance is 50 cm, the error is approximately 24 pixels; for a fixation distance of 280 cm, it reduces to 18 pixels. On the other hand, in the compensated case (same figure, bottom curves) the error remains almost constant when fixation distance changes from 280 cm to 50 cm. These

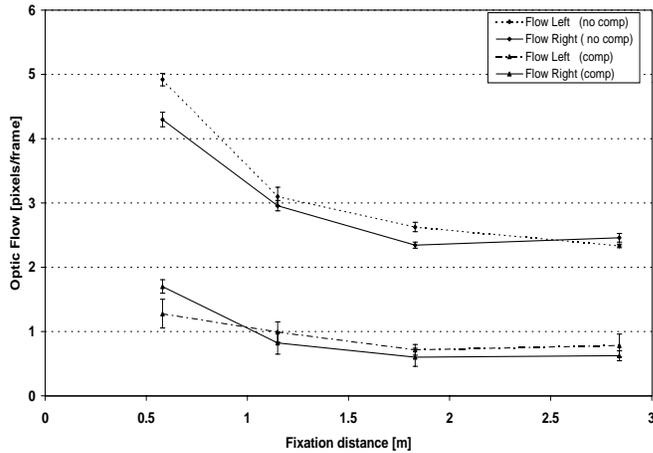


Fig. 15. Optic flow measured during head rotational movements. Upper curves: "non-compensated" case. Lower curves: "adaptive-compensation" case.

results prove that the adaptive tuning of the inertial loop is successful. The values of the adaptive gains derived autonomously by the system on the basis of vergence control, are shown in Fig. 16. The second index of performance, the ROF, is shown in Fig. 15. Note that flow amplitude for fixation distance of 50 cm reduces from 4.5 pixels/frame to about 1.5 pixel/frame in the "compensated case". Furthermore, in the "compensated case" the variability of ROF within the whole fixation range (i.e. 50-280 cm) is smaller than in the non-compensated case. A sequence of frames acquired in the adaptive (left) and non-adaptive (right) stabilized conditions during this experiment is shown in Fig. 17.

5.4 Gaze stabilization during eye-head movements

This section describes the role of inertial information during gaze redirections. In particular, it focuses on fast, coordinated robot eye-head movements generated to foveate targets appearing at eccentric positions. The results show that when the inertial information is used the robot behavior is characterized by a faster fixation and a smaller overshoot of the target.

In a series of repeated trials a target is presented to the robot in a fronto-parallel plane at a distance of 100 cm. Three eccentric positions are chosen, respectively 5, 10 and 15 degrees eccentricity. These eccentricity values represent a reasonable sampling of the robot visual field given the focal length of the cameras (7.5 mm). The movement of the eyes is programmed in retino topic

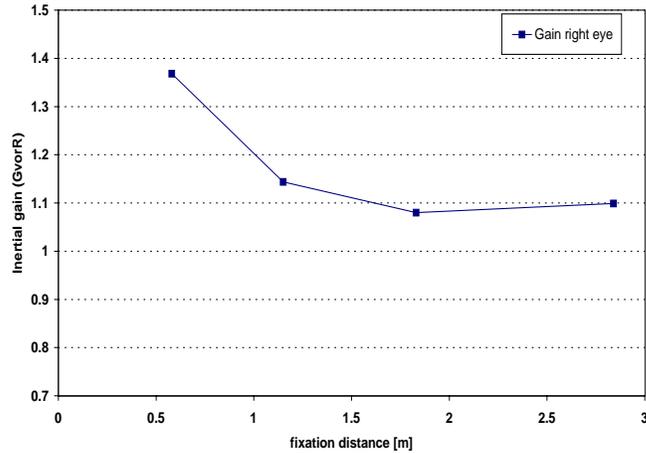


Fig. 16. Adaptive gain G_{vorR} computed autonomously by the system on the basis of control of vergence.

coordinates by transforming image coordinates in motor commands through two look-up tables (LUTs), one for each eye. The look-up tables are filled before the experiment by using a calibration procedure. At the same time a (one-shot) head command is generated taking into account the programmed eye positions at the end of the saccade. Also the head commands are stored in a look-up table. The initial saccadic movement of the eye is open-loop and lasts 4 cycles (the eyes acceleration is set to 50 deg/sec^2 , saccade is performed in $4 \cdot 30 = 120 \text{ ms}$ interval). When the saccadic movement is over, the control of the eye movements is switched to closed-loop. Eye movement is controlled on the basis of color information. The head (i.e. the common pan axis of the cameras) follows the eyes to re-establish a symmetric eye/head configuration, using a PID controller. The error signal driving the common pan axis loop codes the relative orientation of the eyes with respect to the head. The head maximum acceleration is set to 10 deg/sec^2 .

The stabilization performance is measured quantitatively in two cases: a) the inertial information is used by the robot to generate compensatory eye movements, b) the inertial information is not used. The performance measurements are obtained in terms of: 1) the target overshoot, that is, the transients target position error after the saccadic part of the eye control is completed, and 2) the time interval required for ISI to fall below the given threshold. In particular, overshoot is computed as the difference between the minimal retinal error and the maximal retinal error after the saccadic movement. The threshold for the ISI index was experimentally established to be 0.3. In fact, when ISI falls below this value processing of image feature and extraction of dynamic parameters lead to accurate and robust measurement (see optic flow appendix

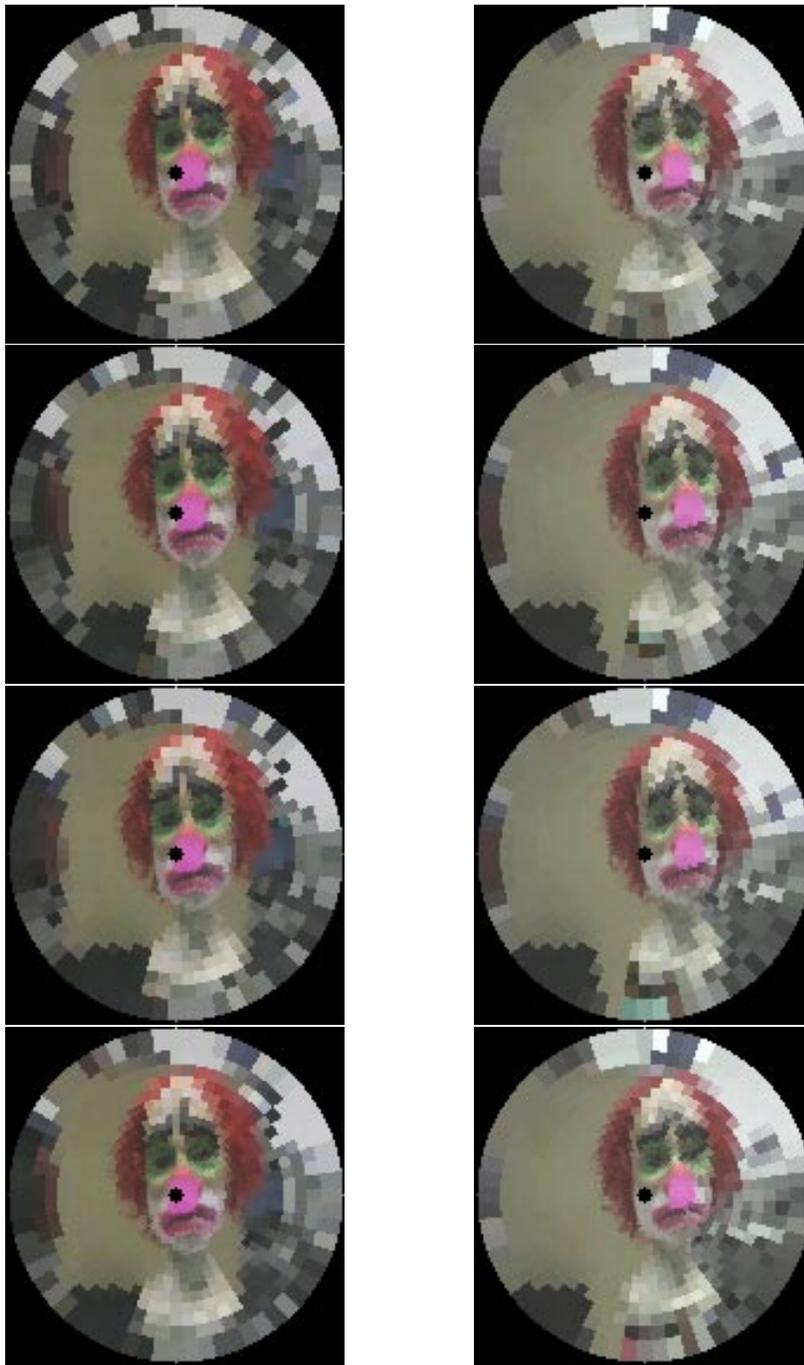


Fig. 17. Left image of stereo pairs acquired during the binocular stabilization experiment. The nose of the clown represents the target to be tracked by color processing. **Left:** images of the adaptive compensated camera. **Right:** images of the non-adaptive compensated camera. In the latter case, note the relative larger error in tracking the target feature. The space-variant images processed by the robot are shown here remapped in Cartesian coordinates.

A.2, range and accuracy measurement).

Figure 18 shows the dynamic trajectories of eye-head coordinated movement

for a saccade performed at 15 degrees angular eccentricity. On the left side, the measurements are obtained when inertial information is used. On the right side, the same measurements are derived when inertial information is not considered. From top to bottom, the trajectories represent: the left eye movement(θ_e), the head movement(θ_h), the angular velocity measured by the inertial system, the corresponding gaze angle ($\mathbf{g} = \theta_h + \theta_e$). Figure 19 shows the target position and the ISI index measured during the same experiment. The comparison of the data in Fig. 19 shows that: i) in the "compensated" case (left), the overshoot of the target is smaller than 12 pixels (see the difference between the marks '*' and '+' around the time unit 15); in the "non-compensated" case the overshoot becomes as large as 35 pixels; ii) the ISI index becomes smaller than the required threshold earlier in time (at about time unit 25 versus 55 in the "non compensated" case). These measurements have been performed several times ($N = 6$) and for different saccade amplitudes, namely 5, 10, 15 degree eccentricity. Data have been averaged and standard deviation computed. Figure 20 summarize the results. The overshoot for the "compensated" case is always smaller than the corresponding "non-compensated" case (see Fig. 19 - left). At the same time, the stabilization interval required for the ISI to fall below the robust threshold in the "compensated" case is considerably smaller. In general the data shows that the use of inertial information in gaze redirection strategies leads to two important advantages: 1) an earlier stable image of the new spotted location and 2) simpler motor control strategies for gaze-line redirection involving coordinated eye-head movements.

6 Discussion

The theoretical and experimental data presented in this paper are aimed at demonstrating the following issues: i) correct/accurate oculo-motor response for gaze stabilization (i.e. eye velocity, G_{vor}) depends on geometrical parameters of the eye-head system and distance of fixation point; ii) the fixation distance has important consequences for both rotational and translational head movements iii) the use of inertial information facilitates the visual processing; iv) an adaptive tuning of the compensatory oculo-motor response can further reduce the visual processing requirements; v) the use of inertial information in gaze redirection strategies increases stabilization performance in dynamic contexts and simplifies coordinated eye-head motor control.

An important point worth stressing, is the difference in latency of the two "sensory" systems. Latencies depend on the amount of sensor-generated data (much larger in the case of visual sensors), to the computational complexity of algorithms used to extract useful sensory cues, and to the available hardware. Inertial data are quantitatively some orders of magnitude less than visual data. As an example, think to robot sampling 3 rotational and 3 linear

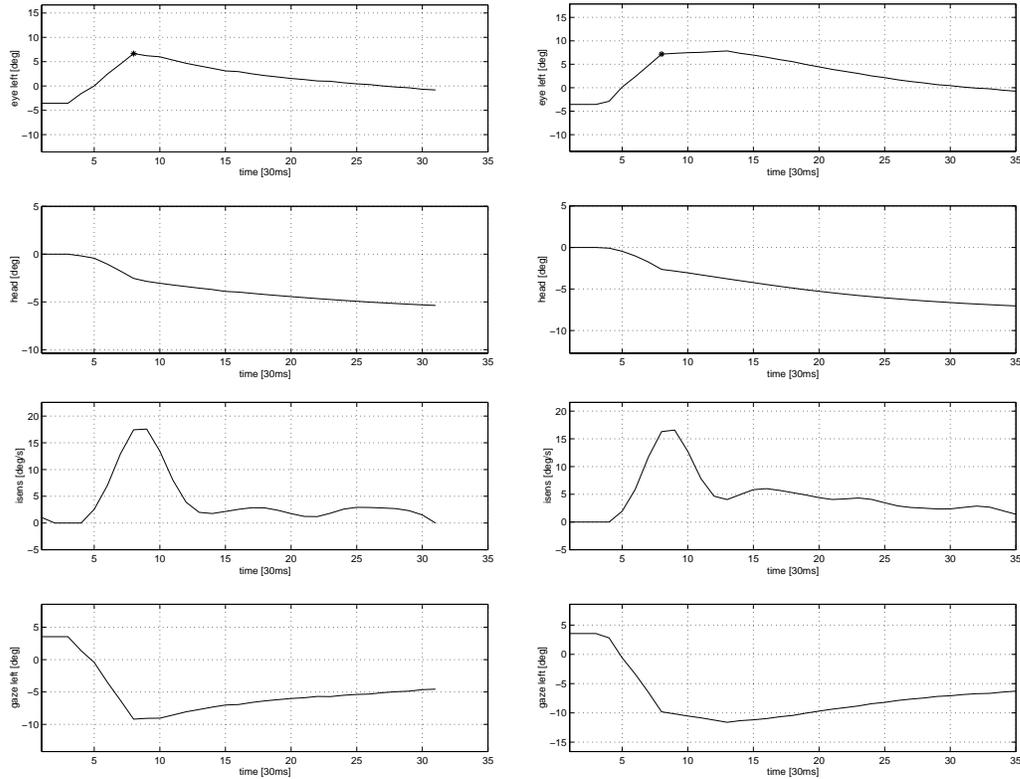


Fig. 18. Position and velocity information during two gaze redirection experiments. From top to bottom: the left eye position, the head position, the inertial sensor output and the gaze position. **Left:** the head velocity sensed by the inertial sensor is used to generate compensatory eye movements. The end of the saccadic part of the eye movement is marked with a '*' symbol. **Right:** inertial information is not used to generate compensatory eye movements.

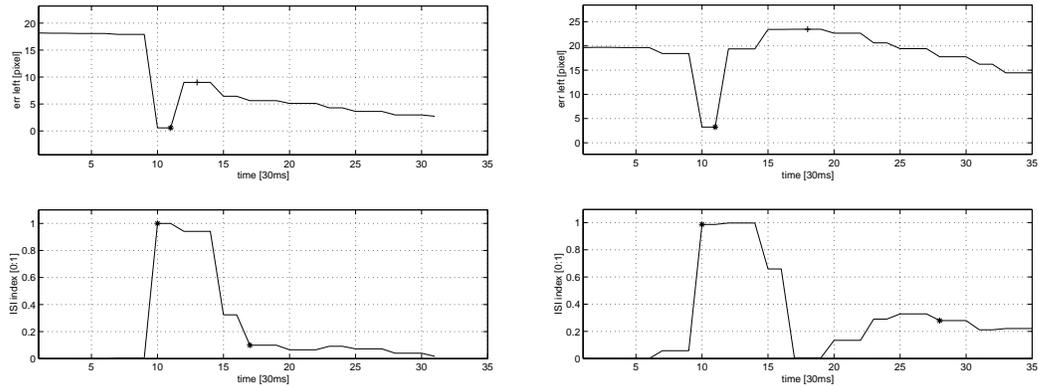


Fig. 19. Visual parameters computed during gaze redirection. **Top:** target retinal error. The overshoot is measured as the difference between the minimum ('*' symbol) and the maximum retinal error ('+' symbol) after the saccade. **Bottom:** image stabilization index (ISI). The time interval required for the ISI to fall below a stable threshold of 0.3 is delimited by two '*' symbols.

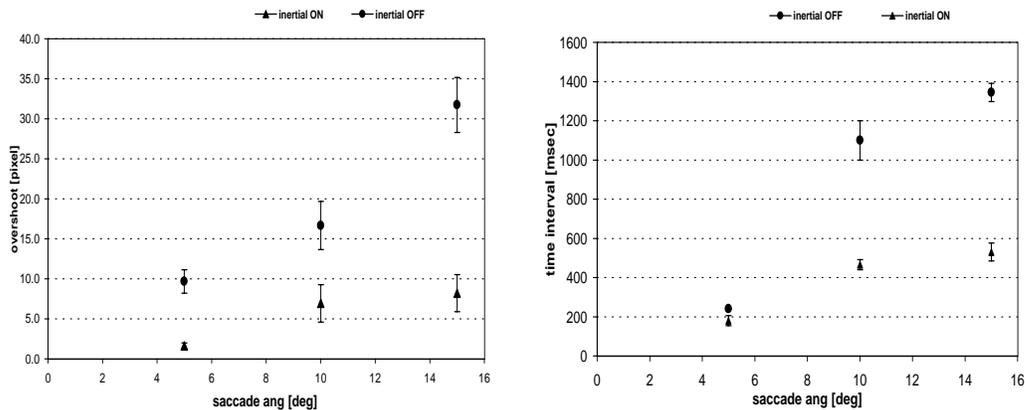


Fig. 20. Parameters measuring stabilization performance during coordinated eye-head movement. **Left:** Retinal error overshoot: the inertial and the non-inertial case. **Right:** time interval required for the ISI to fall below the 0.3 threshold.

accelerometers at 1,000 Hz (which is certainly much more than what the human visual system does) and acquiring and processing stereo images at 25 Hz . The amount of inertial data that needs to be processed per second is almost negligible with respect to the amount of visual data that need to be processed for reliable optical flow estimation. As a direct consequence, gaze stabilization based on visuo-inertial sensory information is more responsive: the number of sensorial sources is increased, but to some extent, the computational resources required are optimized, and as a whole, reduced. Furthermore, the inertial data do not depend on visual processing (and vice versa) and their integration does indeed add a completely new and independent data source. This is particularly true in case of external disturbances where even the proprioceptive information measuring head rotation is not relevant. In this case inertial information seems the only source of extra-retinal signals and a powerful tool to increase the overall performance of the system. As regards the image processing for retinal slip estimation, algorithms adopted for optical flow calculation do indeed have a limited range of image velocities that can be reliably measured. Consequently pure visual stabilization performance is bounded. Whatever this range is, inertial stabilization improves the performance of the system.

An additional important point has been demonstrated in the binocular stabilization experiment. Binocular vision systems can be more efficient in gaze stabilization in at least two senses: i) if a robust control of vergence exists, they can derive consistent depth information and use it to tune the eye compensatory movements. This in turn enables the system to obtain more stable images during translation and rotations at different fixation distances; ii) moreover, a binocular system can exploit binocular constrains (like the horopter) to perform motion analysis limited to the visual area which is interesting to stabilize, that is, the fixation plane (i.e. the plane in space corresponding to intersection of the optical axes, also known as zero disparity plane).

Finally, the experiment on the coordinated eye-head movements has shown the following three main points: first, fixation over-shoot produced by the head movement is considerably reduced. Second, earlier stable fixation of the new spotted location (or target) is obtained. In this regard it is worth noting that algorithms for motion estimation need at least a few stable frames to compute motion parameters consistently. Therefore, the earlier the images are stable, the sooner the "visual functions" of the robot are re-established. This, in turn, increases the robot's reactive behavior to sudden changes occurring at the new spotted location. Third, motor control strategies for coordinated eye-head movements to eccentric targets can be simplified. It is not surprising that biological systems exploit such strategies to increase efficiency of stabilization and to simplify "neural control" of the motor plant during gaze redirection [26]. Further work will concentrate on improving the response of the linear module and on integrating it into the adaptive control scheme.

A Appendix

A.1 Angular velocity of the eye

Consider Fig. 4 where two vectors, namely \vec{v}_g and \vec{v}_b , indicate for the left eye, respectively, the line-of-sight (i.e. the vector connecting the eye position \mathbf{E} with the gaze point \mathbf{P}) and the semi-baseline orientation (i.e. the vector connecting eye position \mathbf{E} and the mid-baseline point \mathbf{B}). The instantaneous eye angle θ_e can be expressed according to the following :

$$\theta_e = \arctan \left(\frac{\vec{v}_b \times \vec{v}_g}{\vec{v}_b \cdot \vec{v}_g} \right). \quad (\text{A.1})$$

Substituting the vector symbols with their expression in terms of \mathbf{a} , \mathbf{b} , \mathbf{d} , θ_h and constraining gaze points to the Z axis, we get:

$$\vec{v}_g = \left(\frac{b}{2} \cos \theta_h - a \sin \theta_h, d - \left(a \cos \theta_h + \frac{b}{2} \sin \theta_h \right) \right) \quad (\text{A.2})$$

$$\vec{v}_b = \left(\frac{b}{2} \cos \theta_h, -\frac{b}{2} \sin \theta_h \right) \quad (\text{A.3})$$

$$\theta_e = \arctan \left(\frac{2(a - d \cos \theta_h)}{b - 2d \sin \theta_h} \right). \quad (\text{A.4})$$

Differentiating eq. A.4 with respect to time and introducing the two auxiliary expressions $Z_l = (a \cos \theta_h + \frac{b}{2} \sin \theta_h)$ and $Z_r = (a \cos \theta_h - \frac{b}{2} \sin \theta_h)$ representing

respectively the Z-coordinates of the left and right eye, we have:

$$\omega_l = \left[\frac{d(d - Z_l)}{d^2 - 2dZ_l + (a^2 + \frac{b^2}{4})} \right] \omega_h \quad (\text{A.5})$$

for the left eye and :

$$\omega_r = \left[\frac{d(d - Z_r)}{d^2 - 2dZ_r + (a^2 + \frac{b^2}{4})} \right] \omega_h \quad (\text{A.6})$$

for the right eye. Equation A.5 can also be rewritten as:

$$\omega_l = \left[1 + \frac{dZ_l - (a^2 + \frac{b^2}{4})}{d^2 - 2dZ_l + (a^2 + \frac{b^2}{4})} \right] \omega_h \quad (\text{A.7})$$

in which the inverse dependence upon distance \mathbf{d} is made more explicit.

A.2 Optic flow: range and accuracy measurements

For the purpose of our measurements a first-order approximation (i.e. affine model) of the optic flow is sufficient and easier to compute than local optic flow [21]. The technique applied to the space-variant image framework is detailed in [9] and revised in [33]. Since we are going to use this motion estimation technique to evaluate the performance of the stabilization control, it is important to know quantitatively the range and the accuracy of the first order optic flow estimates. In particular we are interested in the u_0 horizontal component of optic flow. We have derived this information experimentally by repetitive measurements: a camera fixed on top of a controllable slide is repetitively translated backward and forward at constant velocity while the u_0 estimates are collected. The camera is moving in front of a textured flat surface, 245 cm distant, along a fronto-parallel trajectory. At the end of each back-and-forth motion sequence, the velocity of the slide is increased. The range of translation velocities is between 1 *cm/s* and more than 20 *cm/s*. Figure A.1 shows the u_0 component (left-side) and the Image Stabilization Index (ISI) (right-side) as function of translation velocity. The mean value and the standard deviation are plotted for each experiment. The data show a good behavior of the u_0 estimate for translation velocities up to 17 *cm/s*. For higher translation velocities (more than 20 *cm/s*) the u_0 estimate saturates. These recordings, once again, emphasize in a quantitative way that, for a given amount of computational resources, visual information alone can be used reliably only within the “good range” of measurements, that is, up to the point

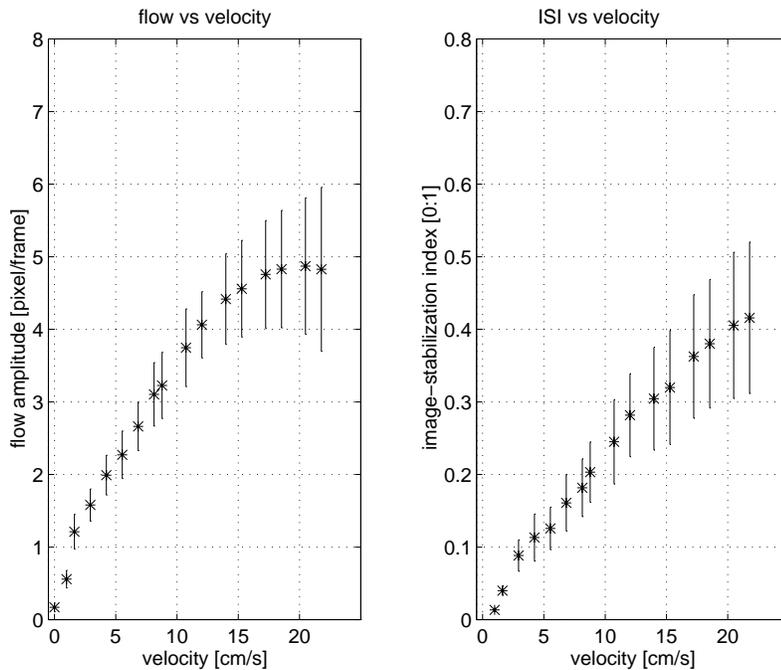


Fig. A.1. Range and accuracy of the optic flow horizontal component u_0 as measured by our algorithm: saturation effect (left), and image stabilization index (right). An ISI value above the 0.3 threshold indicates the beginning of the saturation process in the optic flow measurement.

were a “saturation effect” starts corrupting the accuracy and robustness of the visual measurement process. In the discussion section, this point is further elaborated.

References

- [1] Algrain, M. and Quinn, J. (1993). Accelerometer based line-of-sight stabilization approach for pointing and tracking systems. In *Proc. of Second IEEE Conf. on Control Applications*, Vancouver. IEEE.
- [2] Baarsma, E. and Collewijn, M. (1974). Vestibulo-ocular and optokinetic reactions to rotations and their interaction in the rabbit. *J. Physiology*, 238:603–625.
- [3] Ballard, D. and Brown, C. M. (1992). Principles of animate vision. *Computer Vision Graphics and Image Processing*, 56(1):3–21.
- [4] Barshan, B. and Durrant-Whyte, H. (1995). Inertial sensing for mobile robotics. *Trans. Robotics and Automation*, 11:328–342.
- [5] Benson, A., F.E.Guedry, and Jones, G. M. (1970). Response of semicircular canal dependent units in vestibular nuclei to rotation of a linear acceleration vector without angular acceleration. *J. Physiology*, 210:475–494.

- [6] Biguer, B. and Prablanc, C. (1981). Modulation of the vestibulo-ocular reflex in eye-head orientation as a function of target distance in man. In Fuchs, A. and Becker, W., editors, *Progress in oculomotor research*. Elsevier.
- [7] Brooks, R. (1996). Behavior-based humanoid robotics. In *Proc. IEEE/RSJ IROS'96*, volume 1, pages 1–8.
- [8] Bush, G. A. and Miles, F. A. (1996). Short-latency compensatory eye movements associated with a brief period of free fall. *Exp Brain Res*, 108:337–340.
- [9] Capurro, C., Panerai, F., and Sandini, G. (1997). Dynamic vergence using log-polar images. *International Journal of Computer Vision*, 24(1):79–94.
- [10] Collewijn, H. and Erkelens, C. (1990). Binocular eye movements and the perception of depth. In Kowler, E., editor, *Eye movements and their role in visual and cognitive process: Review of oculomotor research*, pages 213–261. Elsevier, Amsterdam.
- [11] Coombs, D. and Brown, C. (1990). Intelligent gaze control in binocular vision. In *Proc. of the Fifth IEEE International Symposium on Intelligent Control*, Philadelphia, PA.
- [12] Cornilleau-Pérès, V. and Droulez, J. (1994). The visual perception of three-dimensional shape from self-motion and object-motion. *Vision Research*, 34:2331–2336.
- [13] Crane, B. T., Viirre, E. S., and Demer, J. L. (1997). The human horizontal vestibulo-ocular reflex during combined linear and angular acceleration. *Exp Brain Res*, 114:304–320.
- [14] Gellman, R. S., Carl, J. R., and Miles, F. A. (1990). Short latency ocular-following responses in man. *Vis Neurosci*, 5:107–122.
- [15] Gibson, J. (1950). *The perception of the visual world*. Houghton Mifflin, Boston.
- [16] Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Houghton Mifflin, Boston.
- [17] Hine, T. and Thorn, F. (1987). Compensatory eye movements during active head rotation for near targets: effects of imagination, rapid head oscillation and vergence. *Vision Res*, 27:1639–1657.
- [18] Howard, I. (1993). The stability of the visual world. In Miles, F. and Wallman, J., editors, *Visual Motion and its Role in the Stabilization of Gaze*, pages 103–118. Elsevier Science, Amsterdam.
- [19] Kandel, E., Schwartz, J., and Jessel, T. (1991). *Principles of Neuroscience*. Elsevier.
- [20] Keller, E. (1978). Gain of vestibulo-ocular reflex in the monkey at high rotational frequencies. *Vision Research*, 18:311–315.
- [21] Koenderink, J. and van Doorn, J. (1991). Affine structure from motion. *Journal of the Optical Society of America*, 8(2):377–385.

- [22] Lee, D. and Reddish, P. (1981). Plummeting gannets: a paradigm of ecological optics. *Nature*, 293.
- [23] Lobo, J. and Dias, J. (1997). Towards visual and inertial sensing integration: contribution to improve mobile robot autonomy. In *Proc. 5th Symposium on Intelligent Robotics Systems*, Stockholm, Sweden.
- [24] Micheal, J. and Jones, G. M. (1966). Dependence of visual tracking capability upon stimulus predictability. *Vision Research*, 16:707–716.
- [25] Miles, F. A., Kawano, K., and Optican, L. M. (1986). Short-latency ocular following responses of monkey. I. Dependence on temporospatial properties of the visual input. *J Neurophysiol*, 56:1321–1354.
- [26] Morasso, P., Sandini, G., Tagliasco, V., and Zaccaria, R. (1977). Control strategies in the eye-head coordination system. *IEEE Transaction on SMC*, SMC-7 9:639–651.
- [27] Nakayama, K. (1983). Motion parallax sensitivity and space perception. In Hein, A. and Jeannerod, M., editors, *Spatially oriented behavior*, pages 223–241. Springer Verlag, New York.
- [28] Pahlavan, K., Uhlin, T., and Eklundh, J.-O. (1992). Integrating primary ocular processes. In *Proc. ECCV92 - European Conference of Computer Vision*, volume LNCS-588, Santa Margherita Ligure, Italy. Springer-Verlag.
- [29] Paige, G. D. (1983). Vestibuloocular reflex and its interactions with visual following mechanisms in the squirrel monkey. I. Response characteristics in normal animals. *J Neurophysiol*, 49:134–168.
- [30] Paige, G. D. (1991). Linear vestibulo-ocular reflex (lvor) and modulation by vergence. *Acta Otolaryngol Suppl*, 481:282–286.
- [31] Panerai, F. (1995). Inertial sensors for controlled camera systems. Technical Report 2/95, LIRA-Lab, DIST University of Genoa - LIRA lab., University of Genoa.
- [32] Panerai, F. (1998). *Integration of inertial and visual information in binocular vision systems*. PhD Thesis, Department of Communication, Computers and Systems Science, University of Genova, Genova, Italy.
- [33] Panerai, F. and Sandini, G. (1998). Oculo-motor stabilization reflexes: Integration of inertial and visual information. *Neural Networks*, 11(7-8):1191–1204.
- [34] Questa, P. and Sandini, G. (1996). Time to contact computation with a space-variant retina-like c-mos sensor. In *Proc. Int. Conference on Intelligent Robots and Systems*, Osaka - Japan. JRS-IEEE.
- [35] Rogers, B. and Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Res.*, 22:261–270.

- [36] Schwarz, U., Busetini, C., and Miles, F. A. (1989). Ocular responses to linear motion are inversely proportional to viewing distance. *Science*, 245:1394–1396.
- [37] Schwarz, U. and Miles, F. A. (1991). Ocular responses to translation and their dependence on viewing distance. I. Motion of the observer. *J Neurophysiol*, 66:851–864.
- [38] Sharkey, P., Murray, D., Vandeveld, S., Reid, I., and McLauchlan, P. (1993). A modular head/eye platform for real-time reactive vision. *Mechatronics*, 3(4).
- [39] Shelhamer, M., Merfeld, D. M., and Mendoza, J. C. (1995). Effect of vergence on the gain of the linear vestibulo-ocular reflex. *Acta Otolaryngol (Stockh)*, Suppl. 520:72–76.
- [40] Snyder, L. and King, W. (1992). Effects of viewing distance and location of the axis of head rotation on the monkey’s vestibuloocular reflex i. eye movement response. *J. Neurophysiology*, 67:861–874.
- [41] Tabak, S., Collewijn, H., Boumans, L. J., and Van der Steen, J. (1997). Gain and delay of human vestibulo-ocular reflexes to oscillation and steps of the head by a reactive torque helmet. I. Normal subjects. *Acta Otolaryngol*, 117:785–795.
- [42] Telford, L., Seidman, S. H., and Paige, G. D. (1997). Dynamics of squirrel monkey linear vestibuloocular reflex and interactions with fixation distance. *J Neurophysiol*, 78:1775–1790.
- [43] Telford, L., Seidman, S. H., and Paige, G. D. (1998). Canal-otolith interactions in the squirrel monkey vestibulo-ocular reflex and the influence of fixation distance. *Exp Brain Res*, 118:115–125.
- [44] Terzopoulos, D. and Rabie, T. (1997). Animat vision: Active vision in artificial animals. *Videre: Journal of Computer Vision Research*, 1(1):2–19.
- [45] Vieville, T. and Faugeras, O. D. (1990). Computation of inertial information on a robot. In *Proc. of Fifth International Symposium of Robotic Research*, Tokio. MIT Press.
- [46] Viirre, E., Tweed, D., Milner, K., and Vilis, T. (1986). Reexamination of the gain of the vestibuloocular reflex. *J Neurophysiol*, 56:439–450.
- [47] Weiman, C. (1995). Binocular stereo via log-polar retinas. In *Proc. SPIE AeroSense95*, Orlando, Florida.
- [48] Wilson, V. and Jones, G. M. (1979). *Mammalian vestibular physiology*. Plenum Press.