Oculo-Motor Stabilization Reflexes: Integration of Inertial and Visual Information [*]

Francesco Panerai and Giulio Sandini

DIST - University of Genoa

LIRA-Lab (Laboratory for Integrated Advanced Robotics)

Via Opera Pia 13 - 16145 Genoa, Italy

email: nautilus@lira.dist.unige.it

Abstract

Stabilization of gaze is a fundamental requirement of an active visual system for at least two reasons: i) to increase the robustness of dynamic visual measures during observer's motion; ii) to provide a reference with respect to the environment [Ballard and Brown, 1992]. The aim of this paper is to address the former issue by investigating the role of integration of visuo-inertial information in gaze stabilization.

The rationale comes from observations of how the stabilization problem is solved in biological systems and experimental results based on an artificial visual system equipped with space-variant visual sensors and an inertial sensor are presented. In particular the following issues are discussed: i) the relations between eye-head geometry, fixation distance and stabilization performance; ii) the computational requirements of the visuo-inertial stabilization approach compared to a visual stabilization approach; iii) the evaluation of performance of the visuo-inertial strategy in a real-time monocular stabilization task.

Experiments are performed to quantitatively describe the performance of the system with respect to different choices of the principal parameters. The results show that the integrated approach is indeed valuable: it makes use of visual computational resources more efficiently, extends the range of motions or external disturbances the system can effectively deal with, and reduces system complexity.

1 Introduction

One of the key requirements of an active vision system is to be able to keep the direction of gaze stable with respect to static and/or moving points of the environment. This is achieved

either by tracking points electronically in images acquired by static cameras or by moving mechanically the optical axis of the camera (or both). Irrespective of the approach adopted, however, the extraction of image velocity (either by "token tracking", optical flow, or local correlation) is considered a key computational requirement and, as such, has been extensively investigated.

Although in some cases mechanical motions of the cameras may be avoided, a general solution to the problem has to consider motion of the cameras as well as that of the environment (see, for example, [Coombs and Brown, 1990, Pahlavan et al., 1992], [Sharkey et al., 1993, Weiman, 1995]). In most cases, however, the problem of gaze stabilization has been solved relying only on visual information (i.e. visual stabilization approach), that is, by computing image velocity information on specialized hardware, in order to achieve acceptable real-time performance.

On the other hand, considering the different "sources" of image motion, interesting alternatives may be used to reduce the computational requirements of vision based stabilization. In fact, displacements on the image plane may occur due to voluntary motion of the eyes and/or the head, to "external disturbances" (e.g. during locomotion on uneven terrain) or to motion of an object in an otherwise stable environment. In most of these cases, visual information represents just one aspect of the "event" and the use of other sensorial sources, when in appropriate contextual situation, may simplify the general solution. Of particular interest to the present study is the use of proprioceptive information deriving from inertial sensors and the relationship between visual, geometrical and inertial parameters of a camera head for the stabilization of gaze.

Surprisingly, in spite of the fact that most natural systems (at least all those systems equipped with oculo-motor plants) rely not only on visually derived data but also on inertial information to maintain image stability, only few observations have been published in the past on the use of inertial data in artificial systems [Vieville and Faugeras, 1990, Brooks, 1996] and only few implementations have appeared using inertial information for line-of-sight stabilization (e.g. [Algrain and Quinn, 1993]). The goal of this paper is to investigate the role of inertial information in stabilization and gaze-holding and its integration with visually derived data.

The importance of inertial information in image stabilization is evident if, while reading this page, you turn your head to one side. As you do so, your eyes "compensate" by rotating in the opposite direction so that the image of the text remains stable and sharp on your retina. In principle this behavior could be implemented entirely on the basis of visual cues by tracking the part of the image which is currently fixated. Alternatively one could think of using a copy of the motor commands sent to the head, to generate a compensatory motion of the eyes [1]. Finally one could use an inertial sensing device to measure the rotational velocity of the head and exploit this independent source of information (i.e. not relying on vision and/or internal motor commands) to generate compensatory movements. This last solution applies to both voluntary and passive motion of the head and body.

Is the increased complexity implicit in the use of another sensing modality worthwhile? Humans with a defective vestibular systems are so impaired that they are no more able to

---

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

recognize people's faces while walking (and even to read in bed due to heart beat induced head motion). What is the advantage then, from a visual processing point of view? The question is answered through simple mathematical modeling and experimental results, and our conclusion is that the advantage is indeed quite relevant. In particular, the inertial information can be used to limit the range of image velocities required to stabilize the retinal image and, consequently, to reduce the amount of computational resources needed for the stabilization task; the saved resources can be redirected to a different visual task (e.g. tracking, object recognition, binocular fusion), or, they can be used to increase gaze-holding performance.

In biological systems, the mechanism controlling the direction of gaze on the basis of inertial information is called *Vestibulo-Ocular Reflex* (VOR). It's further named *angular* VOR (AVOR) - generating oculo-motor responses to *angular* head motion - and *translational* or *linear* VOR (TVOR or LVOR) - generating responses to *linear* head motion [Schwarz et al., 1989, Paige, 1991]. In the case of the AVOR, the sensing is performed by three semi-circular canals (SCC), which measure angular velocities on three perpendicular directions; in the case of the LVOR, the sensing is performed by the otoliths organs (OTO), which sense linear movements (both in the horizontal and vertical directions) and orientation of the head with respect to gravity [Kandel et al., 1991].

In both cases the inertial information is used to generate motor commands which drive directly the muscles of the eyes to produce compensatory eye movements. A simplified kinematic modeling of gaze stabilization shows that an effective strategy should consider a) the geometric configuration parameters of the eye-head system and b) the geometric relation with the object being fixated. Strictly speaking, if the rotational axis of the eye and head coincides and the fixation point can be assumed at infinity, inertial information alone is sufficient to stabilize gaze perfectly. However it is clear that, in an anthropomorphic binocular eye-head system, the rotation axis of the head and the axes of the eyes are different and, therefore, a rotation of the head causes both a rotation and a translation of the eyes. Consequently compensation for head rotations during gaze-holding cannot be based only on inertial angular measurements, because the compensatory angle depends also on the distance of the object fixated. This residual "error" has to be corrected on the basis of visual information.

This is precisely the main point of the work presented in this paper, i.e. to estimate the range of the residual error and to show that, by integrating inertial and visual measures, the computational requirements are, indeed, constrained without compromising either speed or acceleration of responses.

The paper is organized as follow: section 2 outlines the gaze stabilization framework concentrating on the integration of the visuo-inertial sensory information. Elements of kinematics and modeling of the AVOR for fixation points at close distance are considered. The computational advantages and the wider range of motions/perturbation that the visuo-inertial stabilization approach can deal with are discussed; section 3 describes the robotic apparatus, the performance evaluation criteria, the data collection process; section 4 presents experimental results and leads to final discussion.
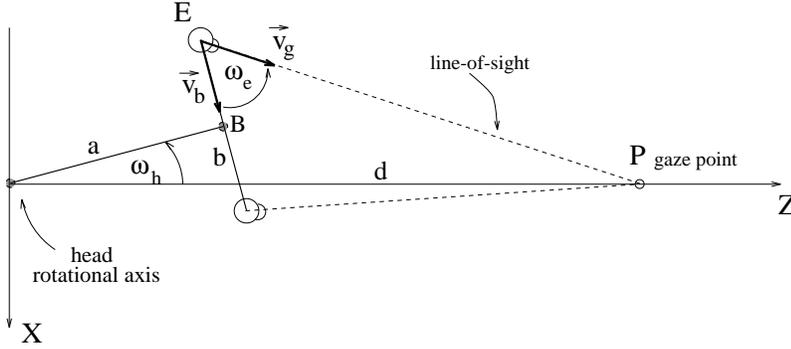
Figure 1: Geometry of the eye-head system showing the parameters relevant to inertial and visual measures.

2 Visuo-inertial stabilization framework

2.1 Kinematics of stabilization

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

In figure 1 the schematic geometry of a binocular system (at least for stabilization around the vertical axis) is shown and the relevant parameters are indicated. In particular, both the inter-ocular distance (or baseline) **b**, and the distance **a** between the rotational axis of the head and the baseline, affect the relationship between eye and head velocity. Moreover the eye velocity required to stabilize the image of a stationary object when the head rotates varies inversely with viewing distance **d**.

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

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

$$\omega_e = \arctan\left(\frac{\vec{v}_b \times \vec{v}_g}{\vec{v}_b \cdot \vec{v}_g}\right). \tag{1}$$

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

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

Differentiating eq. 2 with respect to time gives:

$$\dot{\omega}_e = \left[\frac{d^2 - d\left(\frac{b}{2}\sin\omega_h + a\cos\omega_h\right)}{\frac{b^2}{4} + a^2 + d^2 - b\,d\sin\omega_h - 2a\,d\cos\omega_h}\right] \dot{\omega}_h \tag{3}$$
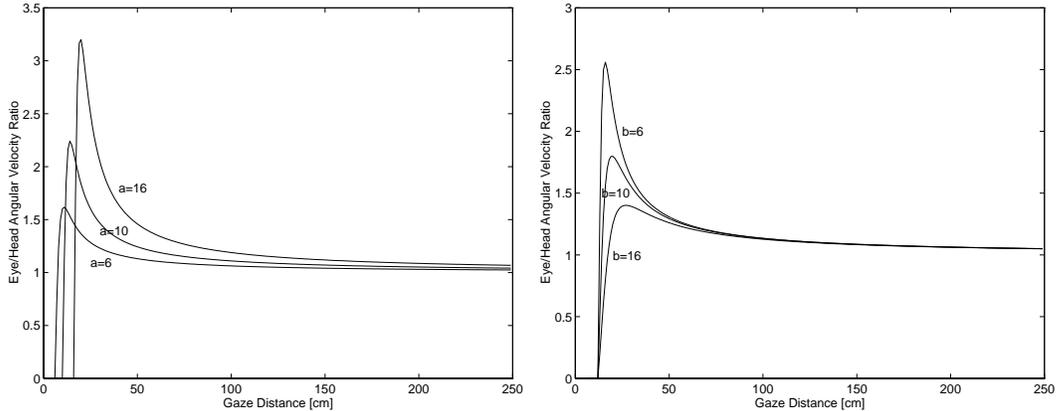
4

Figure 2: Eye-head velocity ratio as a function of distance **d** of the fixation point for different values of **a** and **b** (see geometrical modeling in figure 1). Left: baseline **b** is fixed (6 cm) and eye-to-neck distance **a** increases (from 6 to 16 cm). Right: **a** is fixed (12 cm) while baseline **b** increases (6 to 16 cm). Different eye-head configurations requires different dynamics of the eyes with respect to the head.

Equation 3 determines the relationship between eye velocity $\dot{\omega}_e$ and (a) the geometrical parameters of the eye-head system (i.e **b** and **a**) and (b) distance **d** of the fixation point **P**, for any given head velocity $\dot{\omega}_h$. A more general expression for eccentric gaze points is not considered here but it is easily obtainable (see [Panerai and Sandini, 1997]).

The location of the eyes in the head (defined by **b** and **a**) influences the required eye velocity. In figure 2 the ratio between the eye velocity $\dot{\omega}_e$ and the head rotational velocity $\dot{\omega}_h$ required to stabilize the image of an object, is plotted as function of fixation distance **d**. Different choices of the **b** and **a** parameters determines different shapes of eye-head velocity ratio curves. The sample curves give an idea of possible dynamic requirements of existing active vision systems [Cahan et al., 1992], reflecting different choices of **a** and **b** parameters. With reference to figure 2, it is worth noting that the eye velocity $\dot{\omega}_e$ required to maintain fixation on near objects can be as high as twice the value of $\dot{\omega}_h$. In the [25-200] cm range of fixation, the optimally effective amount of oculo-motor compensation needed to obtain gaze stabilization can change rapidly. Then, an active vision system should consider distance as a tuning parameter for correct oculo-motor compensation.

2.2   Modeling visuo-inertial stabilization

Analysis of visuo-vestibular integration from a bio-engineering point of view can be tracked back to 1977, when Robinson proposed that the ocular stabilization mechanism could be modeled as a mixed inertial feed-forward, visual feedback system [Robinson, 1977]. More recent work of Buizza and Schmid [Buizza and Schmid, 1982] better focused on the characteristics of the visual part of the gaze stabilization system; these models reproduce well the experimental data.

In order to explain from a robotics point of view the relative role of visual and inertial information consider the simple block diagram of figure 3 describing the structure of a

$\dot{\omega}_h$ | Inertial Sensor | $\hat{\dot{\omega}}_h$ | G_{vor}

Vestibular Module

Stabilzation Command | Eyes System | Gaze

retinal slip = 0 | Vision based control
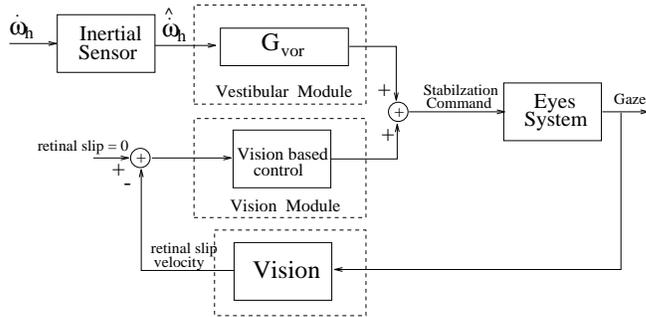
Vision Module

retinal slip velocity | Vision

Figure 3: Block diagram of the control system: the inertial sensory information is processed in open loop acting in a parallel, synergistic way with to the closed loop visual feedback subsystem.

combined visuo-inertial stabilization system.

The performance of the inertial oculo-motor compensation is tuned by the gain (G_{vor}) so that, for example, a value equal to 1 perfectly stabilizes the image of points at infinity; referring to previously kinematic modeling, when distance of fixation **d** becomes much larger than **b** and **a** parameters, the term in square brackets in equation 3 tends to equal unity, and the translation effect induced by rotation is canceled out. Given a configuration of **b** and **a** parameters (e.g. **b** = 6 cm and **a** = 12 cm - approximating the human's) the curves of figure 2 show that for gaze points at close distance, a higher gain is required or, alternatively, a contribution from visually derived measures should be added to the inertial compensation. It is interesting at this point, to consider a quantitative comparison of the computational resources needed in a gaze stabilization task by a system integrating visuo-inertial information and a system using visual information only.

2.3 Residual Ocular Velocity

The contribution of the inertial sensory pathway to oculo-motor compensation can be better understood by examining figure 4. The curves show the eye velocity required to maintain a stable fixation for different constant values of G_{vor} at various distances **d**. The angular velocity of the head considered in this example is $\dot{\omega}_h = 10 \ deg/s$. It is worth noting that with a fixation point at $1 \ m$ a visually-driven compensation of $12 \ deg/s$ is required without inertial stabilization ($G_{vor}=0$) while this value drops to about $2 \ deg/s$ for $G_{vor}=1$. From the visual processing point of view it is precisely this velocity (which we call *Residual Ocular Velocity ROV*) that needs to be measured from image velocity information; the higher the ROV, the more computationally intensive is the visual processing part.

In order to know the amount of visual processing required to compensate the ROV, an estimate of image velocity as function of the geometric parameters of the head-eye system and fixation point distance has to be derived. The equations describing the image velocity due to relative motion between the eye and the scene, as function of eye's motion parameters
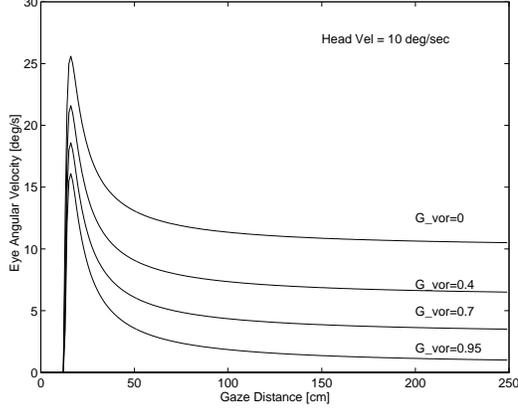
Figure 4: Angular velocity of the left eye for different G_{vor} values. The baseline **b** is set to 6 cm, **a** is equal to 12 cm and $\dot{\omega}_h$ is equal to 10 deg/s.

T_x, T_y, T_z, W_x, W_y, W_z are [Sundareswaran, 1991]:

$$u(x,y) = f_x \left[\frac{\frac{x}{f_x}T_z - T_x}{Z(x,y)} + W_x \frac{xy}{f_x f_y} - W_y(1 + \frac{x^2}{f^2_x}) + W_z \frac{y}{f_y} \right] , \qquad (4)$$

$$v(x,y) = f_y \left[\frac{\frac{y}{f_y}T_z - T_y}{Z(x,y)} + W_x(1 + \frac{y^2}{f^2_y}) - W_x \frac{xy}{f_x f_y} - W_z \frac{x}{f_x} \right] . \qquad (5)$$

To carry out this analysis the horizontal component $u(x,y)$ can be simplified, neglecting the translational and rotational terms T_y, W_x, W_z and considered at the center of the image (i.e. $(x,y) = (0,0)$). These assumptions are based on the hypotheses of considering rotations around the Y axis (i.e. vertical) and translations on the ZX plane only. This yields to:

$$u(0,0) = f_x \left[\frac{-T_x}{Z(x,y)} - W_y \right] . \qquad (6)$$

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

$$W_y = (1 - G_{vor}) \dot{\omega}_h \qquad (7)$$

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

$$T_x = \left(\frac{\mathbf{b}}{2} \cos \omega_e - \mathbf{a} \sin \omega_e \right) \dot{\omega}_h , \qquad (8)$$

equation (6) becomes:

$$u(0,0) = \left[-\frac{\frac{\mathbf{b}}{2} \cos \omega_e - \mathbf{a} \sin \omega_e}{Z(0,0)} - (1 - G_{vor}) \right] \dot{\omega}_h, \qquad (9)$$
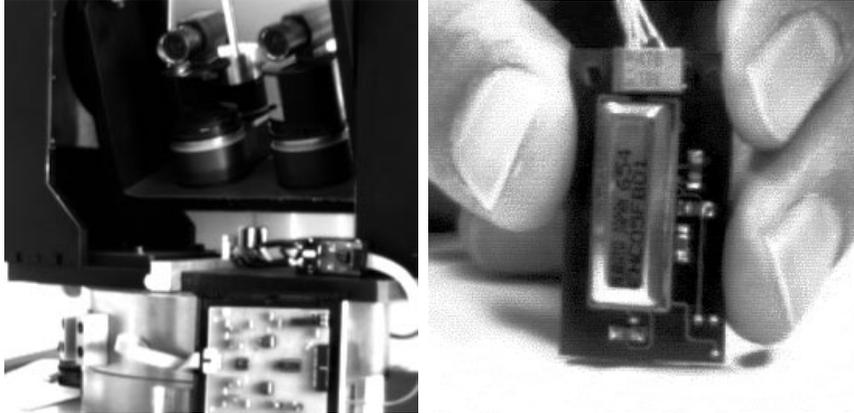
Figure 5: Head mount and head-mounted inertial device.

where $Z(0,0)$ is the actual distance of gaze point $\mathbf{P}$ from eye position $\mathbf{E}$. This relationship can be used to estimate the average image translation u_0 measured in the "center" of the image plane for different values of G_{vor} gain, in response to different head velocities $\dot{\omega}_h$. In the singular case of $G_{vor} = 0$, the equation 9 provides an estimate of the computational requirements of a system using a "visual-only" stabilization approach.

3   Methods and experimental apparatus

The robotic system apparatus is shown in figure 5. It is composed of:

- a 4 d.o.f binocular head [Capurro et al., 1993] equipped with a pair of space-variant cameras, [Ferrari et al., 1995]; only the right camera, though, has been used in the current investigation,

- an angular rate sensor, whose driving and filtering electronics was designed and built in the Lab,

- a Pentium 120 MHz computer and Matrox image processing board.

3.1   Geometry of the camera head and space-variant vision

In order to test the influence of all geometric parameters, the camera is positioned in a stereo arrangement with $\mathbf{b}$=12 cm and $\mathbf{a}$=0.5 cm. A lens with a 4.8 mm focal length is used, covering approximately a range of 50 deg. The camera is connected to a Matrox image processing board simulating a space-variant sensor with about 2000 pixels, arranged in 32 concentric rings, each ring composed of 64 pixels. This type of sensor geometry acquires images in a "log-polar" format (see [Questa and Sandini, 1996] for a detailed description of the sensor's geometry). All visual processing is performed in the log-polar domain. Figure 6-left shows an example of a log-polar image (128x64 pixels); on the right, the corresponding Cartesian reconstructed image is displayed.

Figure 6: The log-polar image geometry on the left (128x64 pixels); the reconstructed Cartesian image on the right.

3.2 Generation of inertial stimulus and sensing device

In order to measure the system performance, a rotational stimulation of the inertial sensor has to be produced. This is achieved by imposing on the head azimuthal axis (i.e. head pan axis) a measurable rotational motion profile. Frequency and amplitude of the oscillatory motion can be varied within the hardware admissible range.

Stabilization of gaze is obtained controlling one degree of freedom: the pan of the right camera. The camera tilt angle is not controlled during the current experiments. The motion profiles of inertial stimulus (i.e. head pan axis) and the eye response are measured reading encoder position.

The inertial sensor used is an angular rate device produced by MURATA (ENC-05EA type), which measures rotational velocity around its longitudinal axis. The output of the sensor is amplified and band-pass filtered (between 0.3 Hz and 5 Hz) by custom designed electronics in the Lab. The lower bound is necessary to eliminate temperature-drifts effects occurring during prolonged operation; the upper bound was chosen in the design phase to limit possible influences of electronic noise in the measuring process [2]. The sampling is performed with a 12 bits A/D converter; the maximum angular velocities measurable by the sensor are, approximately, -90 deg/s and +90 deg/s.

Figure 7 shows, in continuous line, the sensor output for two different motion stimuli (different frequency of head oscillation). The signal, with no additional noise filtering, is used here to generate camera movements with a constant G_{vor} gain of 0.95.

3.3 Measure of stabilization performance

In order to asses quantitatively the performance of the stabilization loop for many different experimental conditions (e.g. different G_{vor} values, different distances), we compute the

---

[2]the MURATA sensor itself has a characteristic band-pass profile, which extends to a maximum of 20 Hz in particular device types.
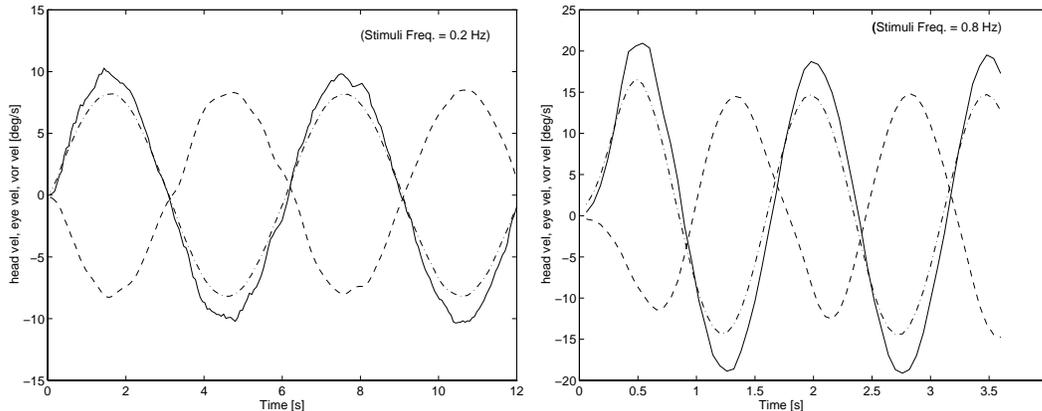
Figure 7: Inertial sensor output for different head stimuli. The curves show the inertial sensor output (continuous line), the head velocity (dash-dot line), and the generated camera movement (dash line). Left: frequency of stimuli 0.2 Hz, peak velocity 10 deg/s. Right: frequency of stimuli 0.8 Hz, peak velocity 15 deg/s.

Image Stabilization Index (ISI) defined in the following way: $ISI_i = 1 - NC_i$, where

$$NC_i = \frac{\sum_{\eta,\xi}(I_i(\eta,\xi) - \mu_i) \cdot (I_{i-1}(\eta,\xi) - \mu_{i-1})}{\sqrt{\sum_{\eta,\xi}(I_i(\eta,\xi) - \mu_i)^2 \cdot \sum_{\eta,\xi}(I_{i-1}(\eta,\xi) - \mu_{i-1})^2}} \tag{10}$$

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

3.4   Measure of image velocity: range and accuracy

For the purpose of our measurements a first-order approximation (i.e. affine model) of the optic flow is sufficient and easier to compute [Koenderink and van Doorn, 1991]. In particular we are interested in measuring the horizontal component in the center of the image; in fact, this estimate is directly proportional to the amount of ROV that can be canceled out by appropriate camera rotation (eye movement).

To compute the values of the affine parameters from log-polar images, the same estimation method already used in the context of dynamic vergence experiments [Capurro et al., 97] is adopted. It is in appendix A, that we revisit the affine model of the optic flow and how image velocity information is extracted in the log-polar domain.

It is important, at this point, to know quantitatively the range and the accuracy of the estimate of u_0 component of the optic flow as measured by our algorithm. We derived this information experimentally by repetitive measurements: a camera fixed on top of a
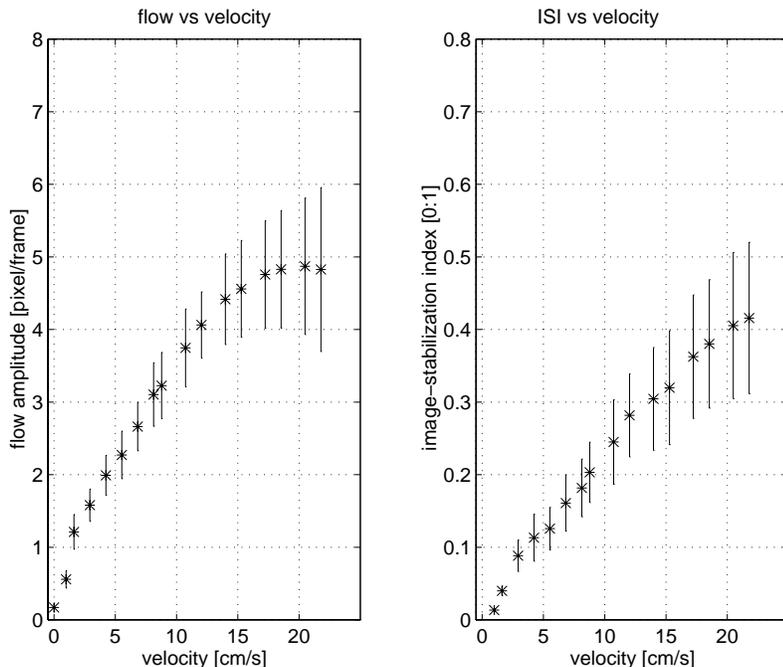
Figure 8: Range and accuracy of the zero-order horizontal flow component u_0 as measured by our algorithm.

controllable slide is repetitively translated backward and forward at constant velocities while the u_0 estimates are collected. The camera is moving in front of a textured flat surface, 245 cm distant, along a front-parallel trajectory. At the end of each back-and-forth motion sequence, the velocity of the slide is increased. The range of translation velocities is between $1\ cm/s$ and more than $20\ cm/s$. Figure 8 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 [3], visual information alone can be used reliably only within the "good range" of measurements, that is, up to the point were a "saturation effect" starts corrupting the accuracy and robustness of the visual measuring process. In the discussion section, this point is further elaborated. We will numerically compare gaze stabilization requirements, in terms of ROV (i.e. residual image velocity), that must be robustly estimated, to generate optimal stabilization, either using visual information, or conversely, using inertial information.

---

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

4 Experimental results

A set of experiments is carried out to quantify the gaze stabilization behavior of the system with respect to different features, and to compare the results with theoretical data. The three experiments envisaged, are intended to characterize: 1) the inertial stabilization alone; 2) the visuo-inertial stabilization; 3) the performance of the system computing dynamic visual cues during visuo-inertial stabilization.

Optic flow and image stabilization index are computed as explained in the methods section. All experiments are performed in real-time with a control period of about 80 msec [4]. During the experiments the following parameters are stored to estimate the overall performance: i) position angle of the inertial stimulus; ii) output of the inertial sensor; iii) u_0 component of optic flow; iv) values of the image stabilization index; v) position angle of the controlled camera. Head and eye velocities are computed off-line on the basis of the sampled angular position using a 5-points approximation of derivative.

4.1 Inertial stabilization experiment

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

A sample output of the data recorded during stabilization is shown in figure 9. With respect to these (and following) figures we would like to point out that the profile of the head velocity is far from sinusoidal. This is a consequence of the way we generated the velocity profile of the inertial stimulation [5].

Figure 9 (upper-left) shows the curves of head velocity, inertial measure and eye velocity for a value of G_{vor}=0.7. The "*" symbol superimposed to the head velocity profile, identifies the relative maxima of head velocity. Figure 9 (upper-right) shows the computed horizontal component of the optical flow u_0. As it can be seen here, the maximum value of the computed optical flow is close to saturation level of the algorithm (compare with results shown in figure 8). In figure 9 (bottom), the stabilization index is shown for the same experiment. The "*" symbols here indicate the values of the ISI corresponding to the maxima of the head velocity profile. These plots are given here to show qualitatively the performance of the inertial stabilization and the performance of the optic flow computation.

In order to quantify these parameters for different values of **d** and G_{vor} the same data are recorded in other experimental sessions and representative numerical values are computed off-line on the acquired measures. This was done by identifying the maxima of the velocity profiles of the oscillatory motions applied to the camera (i.e. the maxima of the velocity

---

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

[5]the servoing of the main head axis for generation sinusoidal trajectories was performed using an external controller. This solution avoided any interference in the stabilization control loop; unfortunately due to some hardware compatibility problem, an optimal tuning of servoing parameters, and consequently, optimal generation of sinusoidal trajectories, was not possible.
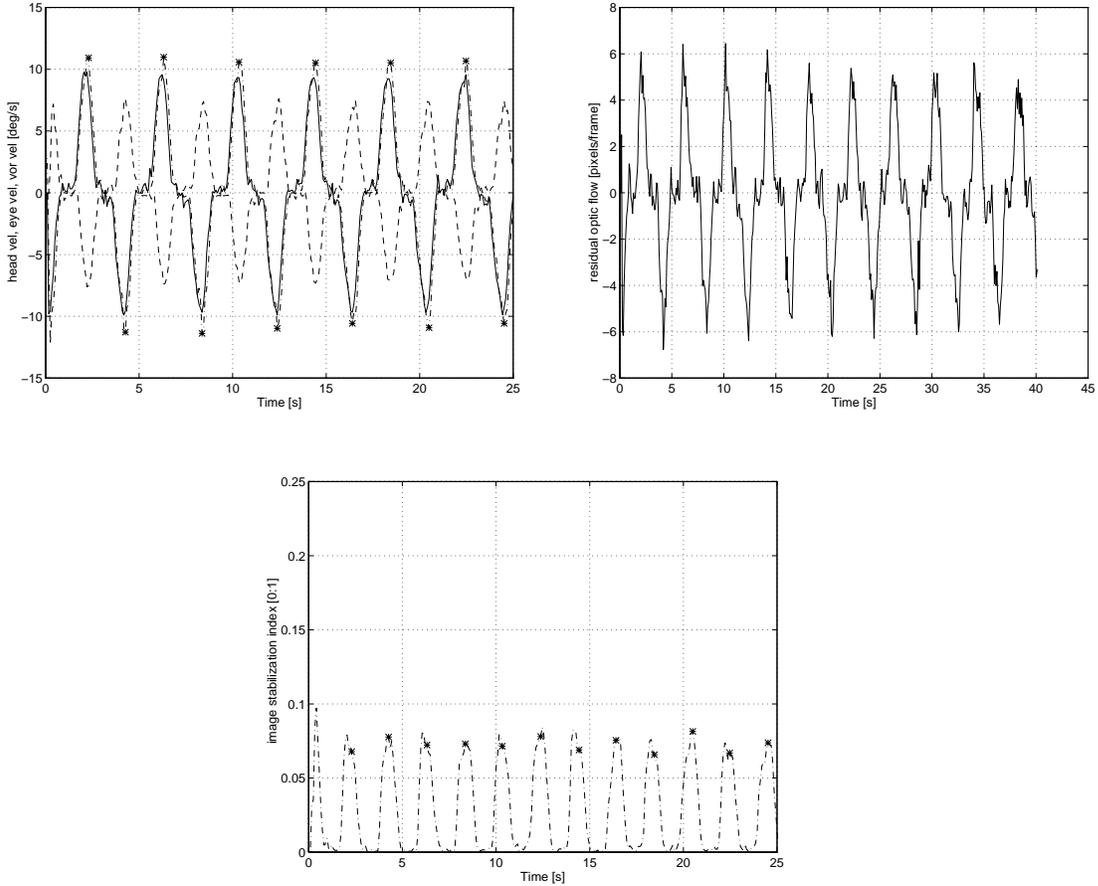
Figure 9: Sample data during one experimental session. (top) Head velocity (dash-dot), inertial sensor output (continuous) and eye velocity (dash). "*" symbols on the head curve identify the maxima of the head velocity profile that are used as time references for optic flow and correlation measurements. In this particular case the value of G_{vor} is equal to .7 and the distance of the fixation point is 240 cm. (bottom) Image stabilization index (ISI) recorded during the same experiment.
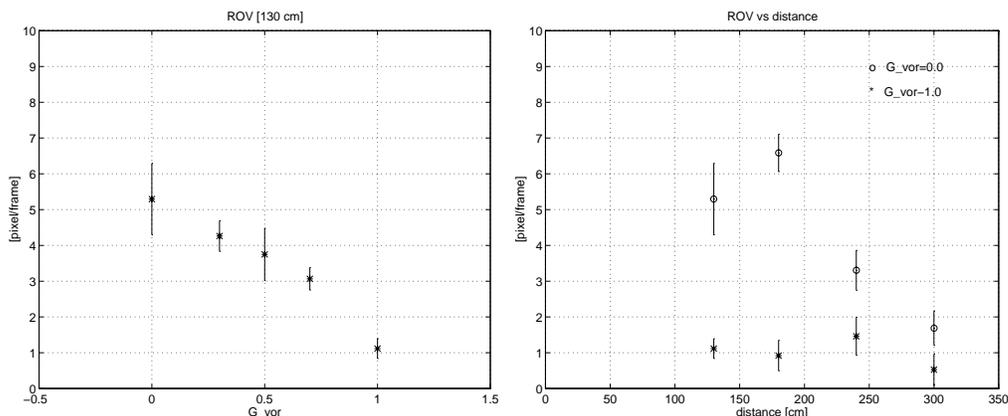
Figure 10: Residual optic flow measured on the image plane. (left) Fixation distance is 130 cm: a decrease of 0.5 in VOR gain, increments by a factor of 3 the ROV. (right) ROV for different distances when two different G_{vor} gains are used.

profiles used to stimulate the inertial sensor). These points are indicated in figure 9 by "*" symbols and are used as time references to extract from the optical flow and ISI data the values to be used as representative data. In all experiments, at least 10 maxima are identified in the head velocity profile, and the mean and standard deviation of the optical flow and correlation at this time instants computed.

The variation of ROV as a function of G_{vor} is shown in figure 10 (left) for a fixation distance of 130 cm. The relationship between G_{vor} and ROV is evident and it is worth noting that, without inertial stabilization (i.e. for $G_{vor}\bar{0}$), even at this relatively long distance ROV approaches the saturation level. If the object gets closer and no inertial loop is present, stabilization based only on vision would be impossible. This is clear in figure 10 (right) where the value of ROV for various distances and two different values of G_{vor} is shown. In fact, while the values obtained for $G_{vor}=1$ (tagged with a '*' symbol) are well maintained within the range of our optical flow estimation, the values at 180 cm and 130 cm for $G_{vor}=0$ (tagged with an 'o' symbol) begin to saturate, giving a wrong estimate. The consequences of this can be seen in figure 11 where the ISI is plotted for different values of G_{vor} and for two different distances of the fixation point (remember that, according to our measure, a high value of the ISI means a bad stabilization).

4.2   Visuo-inertial stabilization experiment

In the second set of experiments the integration of inertial and visual information was implemented. A constant value of G_{vor} is chosen (close to the human VOR gain in darkness) and the residual "retinal-slip" is compensated using visual information. An improved performance is indicated by the image stabilization index (ISI) over different fixation distances. The u_0 component of the affine flow was used to synthesize an oculo-motor command that was linearly added to the inertial information. At the present stage, the control strategy is very simple and does not take into consideration the possibility of different delays in the two control loops (i.e. visual loop and inertial loop). However, even with this simple control
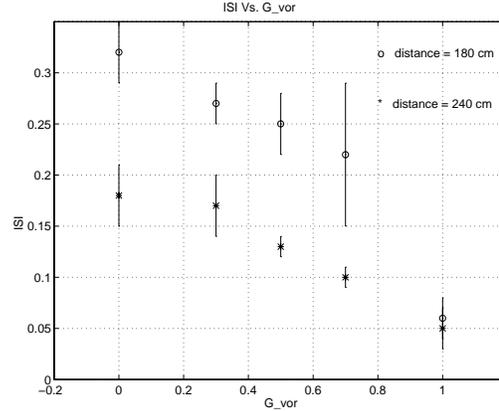
14

Figure 11: Image stabilization index: better stabilization is mapped to lower index values. Stabilization behavior is shown for two different fixation distances (180 symbol 'o', 240 symbol '*').
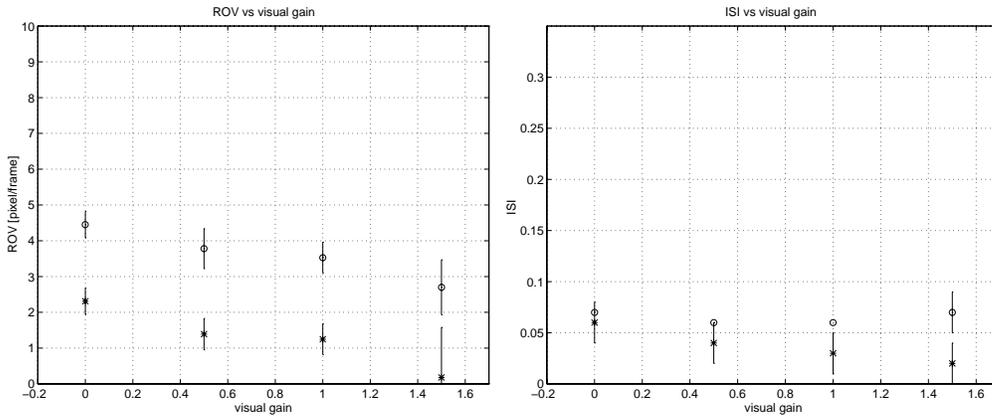

Figure 12: ROV and ISI of two sets of data. G_{vor} is constant within each set, while visual gain is increased; 'o' symbol for $G_{vor} = 0.7$; '*' symbol for $G_{vor} = 1$.

strategy, the advantages obtained using both, visual and inertial information, are evident. The results of these experiments are shown in figure 12. Two sets of residual optic flow and the ISI data are plotted, each set corresponding to a constant G_{vor} value (i.e. $G_{vor} = 0.7$ is represented by 'o' symbol, and $G_{vor} = 1.0$ data represented by '*' symbol). Data within each set are obtained increasing the amount of visual information used to synthesize the final stabilization command (i.e. increasing the visual gain). Performance in stabilization, measured by the image stabilization index (ISI), improves for both values of G_{vor}. As the use of the visual component (i.e u_0) in the control loop increases (i.e. by increasing the vision gain), the resulting ISI becomes smaller.

In figure 12 (left) the residual optic flow (ROV) for visual gain = 0.0 (i.e. only inertial information is used in the control loop) is estimated in 2.3 pixels/frame for G_{vor}=1.0 and 4.5 pixels/frame for G_{vor}=0.7. This values reduce almost to zero pixels/frame and to less than 2.7 pixels/frame when visual information is introduced with a gain = 1.5. The resulting

Figure 13: Binocular frames during stabilization experiment. Left: non stabilized camera; right: stabilized camera.
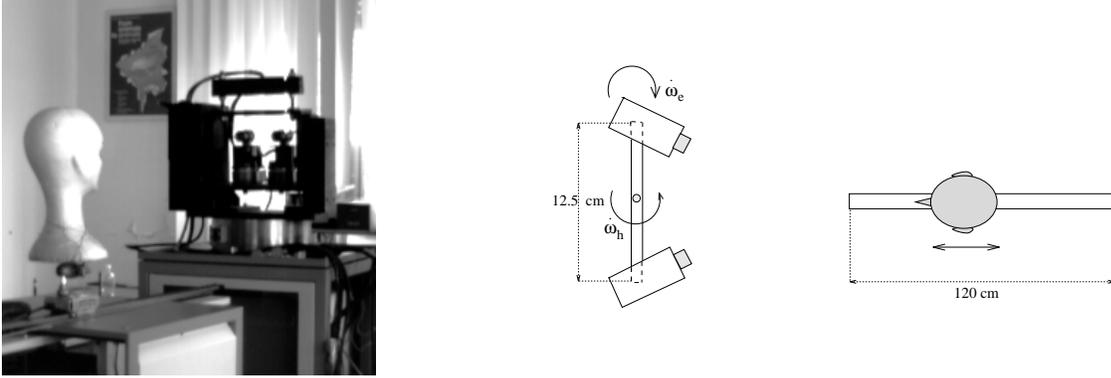
Figure 14: Setup used for estimating divergence during oscillatory head motion.
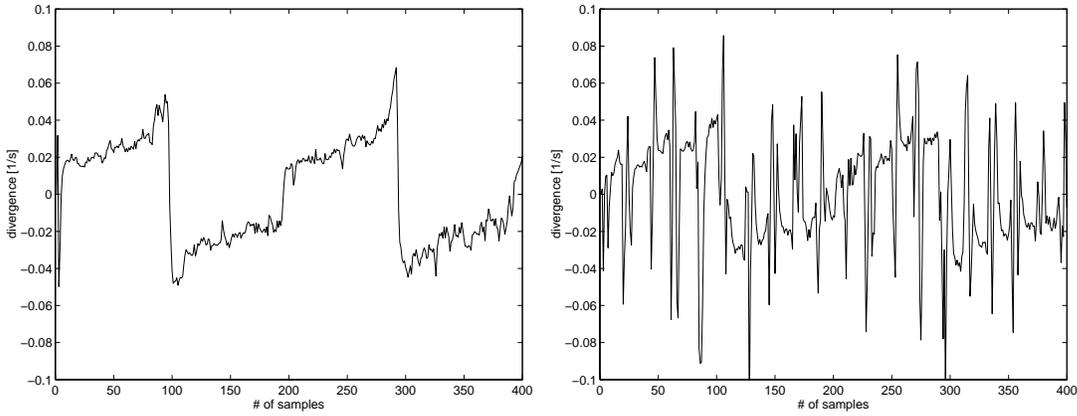

Figure 15: Divergence estimated in stabilized (left) and non-stabilized (right) operating conditions.

improvement in stabilization is also evident in the ISI plots (see figure 12 (right)), where the ISI linearly decreases, thus indicating better overall image stability. Figure 13 shows in binocular frames, the simultaneous left camera view (non-stabilized) and the right camera view (stabilized) of the vision system, during a sinusoidal oscillatory head motion [6].

4.3 Dynamic visual measurement during gaze stabilization

The visuo-inertial stabilization strategy is tested in relation to a dynamic visual measurement task, by fixating an approaching object while the head is oscillated sinusoidally at a given amplitude and frequency (see figure 14). In this situation the expansion/contraction pattern produced by object motion on the image plane is estimated by computing divergence (see appendix A for details on divergence measurement), a positive divergence indicating that the object is approaching, a negative divergence indicating the object is moving farther away. Figure 15 (left) shows the estimate of divergence during visuo-inertial stabilization. The profile of divergence clearly alternates between positive (object approaching) and neg-

---

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

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

Table 1: Image velocities for gaze point at distance $\mathbf{d} = 100$ cm. Frame rate is 12.5 Hz (corresponding to our experimental 80 msec control cycle). Focal length value measured in pixels is approximately 260.

ative values (object moving away). It is in these operative conditions, that the estimated divergence can be used successfully in a control loop (e.g. to control vergence dynamically [Capurro et al., 97]).

On the contrary, figure 15 (right) shows divergence estimate in non-stabilized gaze condition. The clear alternation between positive and negative "behavior" disappears and use of this visual information in a control algorithm will be inappropriate.

5 Discussion

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

In section 2 an expression of the amount of residual ocular velocity (measured on the image plane) for different values of G_{vor} gains has been presented. In table 1 this expression is used to derive estimates of the average image velocities on the image plane in the following two cases: gaze stabilization is performed using only visual information ($G_{vor} = 0$) and gaze stabilization is performed using inertial information (G_{vor}= 0.9). In the first case, the table shows that the range of motion/external disturbances the system can effectively deal with, is limited by the visual measurement process [7]. A saturation effect will cause insufficient oculo-motor compensation and consequently inefficient stabilization even for non-high values of inertial stimulation (see table 1 for $\dot{\omega}_h = 20 \ deg/s$). The integrated approach, in contrast, seems much more effective and better suited to deal with a wider range of motion or disturbances, overcoming the limitations of the visual stabilization.

The data presented in the experimental results are, of course, dependent on the hardware and software components used, and particularly, on the kind of sensors and optical flow esti-

---

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

mation algorithm. However, the constraints introduced by these choices are indeed general, and we belive that, the results obtained in this paper could be generalized to other robotic visual systems. For example, it is conceivable that all algorithms adopted for optical flow estimation do indeed have a limited range of image velocities that can be reliably measured and consequently image stabilization performance is also bounded. Whatever is this range, inertial stabilization improves the performance of the system.

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

The limitations of the present work are, on one side, the use of just one inertial sensor (and consequently the stabilization is limited to one degree of freedom), on the other, the rather simple approach adopted to integrate visual and inertial information and the inappropriate computer architecture, not allowing any multi-tasking, multi-rate visuo-inertial servoing. These issues deserve further investigation. One last point worth mentioning is the extension to a binocular system and the possible use of disparity and proprioceptive eye-angle information to tune the gain of the inertial loop. This aspect is currently investigated and may prove even more interesting than simply integrating inertial and optical flow information.

A Appendix

A.1 First-order optic flow estimation method

The affine model of the optic flow [Koenderink and van Doorn, 1991] can be described on the basis of four quantities: image *translation*, *rotation*, *divergence* and *shear*. The first two terms specify respectively a rigid two-dimensional translation and rotation of the fixated object. The third term describes an isotropic expansion (or contraction) that specifies a change in scale or a pure deformation. The shear term results in a distortion of the image pattern and corresponds to an expansion in a specified direction and a simultaneous contraction in the

perpendicular one in such a way that the area of the pattern is preserved. The component of the affine motion relevant for our measurements is *translation*. The first-order approximation of optic flow (u, v) can be written as:

$$\begin{bmatrix} u \\ v \end{bmatrix} \cong \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} D + S_1 & S_2 - R \\ R + S_2 & D - S_1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \tag{11}$$

where u_0 and v_0 are the uniform, zero-order components of the flow at the center of the image, D is *dilation*, R, *rotation*, and S_1 and S_2 are the components of *shear*. The equation (11) can be re-arranged by taking into consideration the Cartesian to log-polar image velocity transformation and the brightness constancy assumption [Horn, 1986], thus resulting in:

$$-\dot{E} = \begin{bmatrix} \gamma_1 & \gamma_2 & g_\xi & g_\eta & \gamma_3 & \gamma_4 \end{bmatrix} \cdot \begin{bmatrix} u_0 & v_0 & D & R & S_1 & S_2 \end{bmatrix}^T \tag{12}$$

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

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

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

Explicit knowledge of partial derivatives $\frac{\partial E}{\partial \eta}$, $\frac{\partial E}{\partial \xi}$ and temporal derivative $\frac{\partial E}{\partial t}$ in at least six points of the image is required to solve the system (12) using a least square method.

References

[Algrain and Quinn, 1993] 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.

[Ballard and Brown, 1992] Ballard, D. and Brown, C. M. (1992). Principles of animate vision. *CVGIP*, 56(1):3–21.

[Brooks, 1996] Brooks, R. (1996). Behavior-based humanoid robotics. In *Proc. IEEE/RSJ IROS'96*, volume 1, pages 1–8.

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

[Cahan et al., 1992] Cahan, U., von Seelen, and Madden, B. (1992). Binocular camera plat-
form home page. Technical Report
http://www.cis.upenn.edu/ grasp/head/headpage/headpage.html, GRASP Lab - Univ.
of Pensylvania, Philadelpia, USA.

[Capurro et al., 1993] Capurro, C., Panerai, F., and Grosso, E. (1993). The lira-lab head.
Technical Report LIRA-Lab TR-3-93, University of Genova, Genova, Italy.

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

[Coombs and Brown, 1990] Coombs, D. and Brown, C. (1990). Intelligent gaze control in
binocular vision. In *Proc. of the Fifth IEEE International Symposium on Intelligent Con-
trol*, Philadelphia, PA.

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

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

[Kandel et al., 1991] Kandel, E., Schwartz, J., and Jessel, T. (1991). *Principles of Neuro-
science*. Elsevier.

[Koenderink and van Doorn, 1991] Koenderink, J. and van Doorn, J. (1991). Affine struc-
ture from motion. *Journal of the Optical Society of America*, 8(2):377–385.

[Pahlavan et al., 1992] 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.

[Paige, 1991] Paige, G. D. (1991). Linear vestibulo-ocular reflex (lvor) and modulation by
vergence. *Acta Otolaryngol Suppl*, 481:282–286.

[Panerai and Sandini, 1997] Panerai, F. and Sandini, G. (1997). Integration of inertial and
visual information in an active vision system. Technical Report 4/97, LIRA-Lab, Univer-
sity of Genoa, Genova, Italy.

[Questa and Sandini, 1996] 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.

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

[Schwarz et al., 1989] Schwarz, U., Busettini, C., and Miles, F. A. (1989). Ocular responses
to linear motion are inversely proportional to viewing distance. *Science*, 245:1394–1396.

[Sharkey et al., 1993] Sharkey, P., Murray, D., Vandevelde, S., Reid, I., and McLauchlan, P. (1993). A modular head/eye platform for real-time reactive vision. *Mechatronics*, 3(4).

[Sundareswaran, 1991] Sundareswaran, V. (1991). Egomotion from global flow field data. In *Proc. of the IEEE Workshop on Visual Motion*, Princeton, NJ - USA.

[Vieville and Faugeras, 1990] 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.

[Weiman, 1995] Weiman, C. (1995). Binocular stereo via log-polar retinas. In *Proc. SPIE AeroSense95*, Orlando, Florida.