Proceedings of the 2008 IEEE
International Conference on Robotics and Biomimetics
Bangkok, Thailand, February 21 - 26, 2009

Robotic Implementation of Biological Bayesian Models Towards Visuo-inertial Image Stabilization and Gaze Control *

Jorge Lobo, João Filipe Ferreira and Jorge Dias
ISR FCT-UC University of Coimbra
3030-290 Coimbra, Portugal
{*jlobo, jfilipe, jorge*}*@isr.uc.pt*

***Abstract*—Robotic implementations of gaze control and image stabilization have been previously proposed, that rely on fusing inertial and visual sensing modalities. They are bioinspired in the sense that human and biological system also combine the two sensing modalities for the same goal. In this work we build upon these previous results and, with the contribution of psychophysical studies, attempt a more biomimetic approach to the robotic implementation. Since Bayesian models have been successfully used to explain psychophysical experimental findings, we propose a robotic implementation using Bayesian inference.**

I. Introduction

It is well known that the information provided by the vestibular system is used during the execution of visual movements such as gaze holding and tracking [1]. Neural interactions of human vision and vestibular system occur at a very early processing stage [2]. Human sense of motion is derived from two main factors: the contribution of the vestibular system and retinal visual flow. The inertial information enhances the performance of the vision system, and the visual cues aid the spatial orientation and body equilibrium.

Artificial systems should also exploit this sensor fusion [3]. Micromachining enabled the development of low-cost single chip inertial sensors that can be easily incorporated alongside the camera's imaging sensor, providing an artificial vestibular system. Inertial sensors coupled to cameras provide valuable data about camera ego-motion and how world features are expected to be oriented [4]. Object recognition and tracking benefits from both static and inertial information.

Perception has been regarded as a computational process of unconscious, probabilistic inference. Aided by developments in statistics and artificial intelligence, researchers have begun to apply the concepts of probability theory rigorously to problems in biological perception and action. One striking observation from this work is the myriad ways in which human observers behave as near-optimal Bayesian observers, which has fundamental implications for neuroscience, particularly in how we conceive of neural computations and the nature of neural representations of perceptual variables [5].

*This work is partially supported by EC-contract BACS FP6-IST-027140.

The fact that there is strong evidence for a probabilistic computational framework in the human brain for perception, also brings forth the notion of *optimal percept*, or, in other words, that our percepts are our best guess as to what is in the world, given both sensory data and prior experience [6], [7]. Such an "optimal guess" based on priors also suggests an explanation to why biological perception systems, when faced with perceptual scenarios which do not comply to the statistics of natural environments or when impaired due to disease or cerebral lesions, often fail to perceive the world as it is, substituting its correct description by the erroneous percepts called *perceptual illusions* — these are a direct result of perceptual ill-posed problems [8]. On the other hand, at present, the performance of the human perceptual system is superior in almost every respect to that of machine perception systems. Due to this fact, a significant amount of current research on artificial perception is shifting towards a *bioinspired approach*, implying a Bayesian framework for artificial perception models.

A. Related Work

Comparison of camera rotation estimate given by image optical flow with output from a low cost gyroscope was done for gaze stabilisation of a rotating camera [9]. In [10] the integration of inertial and visual information in binocular vision systems was studied. The image stabilization proposed in [11] relies on a self-tuning neural network that combines inertial and optic flow motion cues and generates oculo-motor compensatory behaviours that stabilize the visual scene. The learning scheme adapts the neural network parameters after a short training period. Critical performance issues are the delays in the visual and inertial loops, as well as the saturation of optical flow measures in the simplified model used.

More recently, a high speed gaze control system based on the Vestibulo-Ocular Reflex has been proposed [12]. The gaze controller involves a feedback control system based on the retinal position error measurement and a feedforward control system based on the angular head velocity measurement. The performance in terms of head rotational disturbance rejection is comparable to that afforded by the human vestibulo-oculomotor system. Visual fixation has also

been suggested as a mechanism for image stabilization [13].

To better exploit the benefits of combining the two sensing modalities in artificial systems, a clear understanding of biological systems is important. Vestibular information is necessary not only for vestibular reflexes but also in various cognitive functions for our adequate behaviour in three-dimensional space. In [14] the regions of the cerebral cortex where vestibular information is represented is investigated. Perception and action influence each other [15], making some biological systems highly coupled and complex, from which direct models for sensor fusion are not easily derived. In [16], [17] the role of gravity in visual perception and how the brain deals with the ambiguity between inclination and body acceleration is investigated. In [18], [19] the motion perception inferred from visuo-vestibular cues is studied. The perceived relative motion is important for posture control [20]. Taking advantage of improved brain imaging techniques, a better understanding of the visual motion and self-movement interactions has been pursued [21], [22]. A Bayesian model of human processing of vestibular information is proposed by [23], obtaining satisfactory responses to complex motion stimuli.

II. Image Stabilization and Gaze Control

Our robotic implementation of gaze control and image stabilization will basically perform visuo-inertial servoing, exploring the complementarity of the two sensing modalities, but the sensory inputs will be processed using Bayesian inference.

The image optical flow provides motion data from the visual sensor that complements the inertial data.

Taking as a reference the visual servo control topology presented in [24], fig. 1 presents the system block diagram for our robotic system.

A. Probabilistic Block Matching Optical Flow

Recent findings regarding human perception strongly suggest that the brain codes even complex patterns of sensory uncertainty in its internal representations and computations

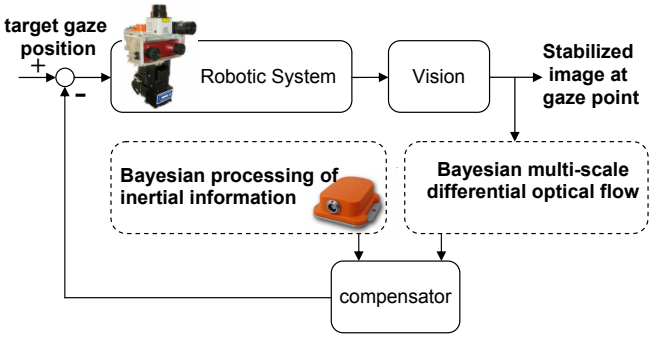

Fig. 1. System block diagram for our robotic system implementation of gaze control and image stabilization.

An example of a common neuronal representation of uncertainty in the human brain is believed to be *neural population coding* (e.g., average firing rate) [5], [25].

With this in mind, we set off in implementing a simple probabilistic optical flow algorithm that would encompass the notion of a population code-type data structure storing two-dimensional *pdf*s on the image velocity space $(\Delta u, \Delta v)$ as an output. The immediate advantage of such an implementation would be the availability of confidence measures per local image velocity measurement taken from the covariance matrix of the corresponding *pdf*. A further advantage would be the possibility of future use of powerful probabilistic/belief propagation methods to allow for temporal integration of several frames. Finally, the probabilistic nature of such an algorithm will allow effortless integration with higher-level Bayesian cue integration modules.

The algorithm is primarily based on Zelek's [26] adaptation of the block matching (correlation) algorithm presented in [27].

A smooth motion field is assumed and also a Gaussian prior probability density in which slower speeds are more likely. The output of the algorithm is a Gaussian distribution over the space of image velocities, at each position in the image. The mean of the distribution is a gain-controlled modification of the basic optical flow solution. The covariance matrix captures directional uncertainties, allowing proper combination with inertial data.

B. Bayesian processing of inertial data

To process the inertial data, we follow the Bayesian model proposed by Laurens and Droulez [23], adapted here to the use of inertial sensors instead of the vestibular system. The aim here is to provide an estimate for the current angular position and angular velocity of the system, that mimics the human vestibular perception.

To overcome the non-linearity of the motion equations and the high dimension space of possible distributions, particle filtering is used. The fact that some motions are more probable than others in human head motion is also replicated in the robotic version, limiting periods of sustained acceleration and also long duration rotations at constant velocity.

1) Motion variables: In this model, X, Y and Z refer to the three axes of the robotic vision head in egocentric coordinates. The orientation of the system in space is encoded using a rotation matrix $\mathbf{\Theta}$. Angular velocity of the head is encoded using the yaw y, pitch p and roll r conventions. Yaw rotations are rotations around the Z axis; pitch around the Y axis and roll around X. When a rotation consists of a combination of yaw, pitch and roll rotation, the three rotations are applied successively and in this order. The rotation update is given by

$$\mathbf{\Theta}^{t+\delta t} = \mathbf{\Theta}^{t}.\boldsymbol{R}(\delta y, \delta p, \delta r) \quad (1)$$

where $\boldsymbol{R}(y,p,r) =$

$$\begin{bmatrix} c(y).c(p) & c(y).s(p).s(r) - s(y).c(r) & c(y).s(p).c(r) + s(y).s(r) \\ s(y).c(p) & c(y).c(r) + s(y).s(p).s(r) & -c(y).s(r) + s(y).s(p).c(r) \\ -s(p) & c(p).s(r) & c(p).c(r) \end{bmatrix}$$

where $c()$ and $s()$ are short form for $\cos()$ and $\sin()$.

The instantaneous angular velocity is defined as the vector:

$$\boldsymbol{\Omega} = \begin{pmatrix} \delta y/\delta t \\ \delta p/\delta t \\ \delta r/\delta t \end{pmatrix}$$

Linear motion of the head is described by the position of the centre of the head in a geocentric reference frame, defined as a position vector $\boldsymbol{M}$. The linear acceleration $\boldsymbol{A}$ is the second derivative of $\boldsymbol{M}$ over time. In our case we are only concerned with the linear acceleration, since gravity will provide an absolute reference for orientation only when $\boldsymbol{A} = 0$.

The sate of our system at time t is therefore defined by $(\boldsymbol{\Theta}^t, \boldsymbol{\Omega}^t, \boldsymbol{A}^t, \boldsymbol{F}^t)$.

2) Sensory input: The calibrated inertial sensors in the Inertial Measurement Unit (IMU) provide direct egocentric measurements of body angular velocity and linear acceleration (including gravity $\boldsymbol{G}$). Given the motion of the system, we can define the probability distribution of the sensory inputs.

The gyros will measure $\boldsymbol{\Omega}^t$ with added Gaussian noise, i.e. $\boldsymbol{\Phi}^t = \boldsymbol{\Omega}^t + \eta_\Phi^t$, where η_Φ^t is a three-dimensional vector, the elements of which follow independent Gaussian distributions with mean 0 and standard deviation σ_Φ.

The accelerometers will measure the gravito-inertial acceleration $\boldsymbol{F}$ with added Gaussian noise, i.e. $\boldsymbol{\Upsilon}^t = \boldsymbol{F}^t + \eta_\Upsilon^t$, where η_Υ^t is a three-dimensional vector, the elements of which follow independent Gaussian distributions with mean 0 and standard deviation σ_Υ. $\boldsymbol{F}$ is the resultant acceleration due to linear acceleration and gravity. Given the geocentric body linear acceleration $\boldsymbol{A}$ and the system orientation $\boldsymbol{\Theta}$, we can compute $\boldsymbol{F}$. In a geocentric frame of reference gravity is a vector $\boldsymbol{G} = (0, 0, -9.81)$, and the gravito-inertial acceleration is given by $\boldsymbol{G} - \boldsymbol{A}$, transforming to the egocentric frame of reference we have

$$\boldsymbol{F} = \boldsymbol{\Theta}^{-1}.(\boldsymbol{G} - \boldsymbol{A}) \tag{2}$$

The sensor data at time t is therefore defined by $(\boldsymbol{\Phi}^t, \boldsymbol{\Upsilon}^t)$.

3) A priori: As suggested in [23], even in the absence of any sensory information, motion estimates for which the rotational velocity and acceleration are low are more probable. This can be described in a simple way using a Gaussian distribution. Having

$$\mathcal{N}_{x,\mu,\sigma} = \frac{e^{-(x-\mu)^2/(2.\sigma^2)}}{\sqrt{2.\pi.\sigma^2}}$$

the probability distribution for acceleration is given by $P(\boldsymbol{A}^t) \propto \mathcal{N}_{|\boldsymbol{A}^t|,0,\sigma_A}$. Similarly for angular velocity $\boldsymbol{\Omega}$ we have $P(\boldsymbol{\Omega}^t) \propto \mathcal{N}_{|\boldsymbol{\Omega}^t|,0,\sigma_\Omega}$.

Program
- Description
  - Specification
    - Relevant variables:
      - $\boldsymbol{\xi}^t = (\boldsymbol{\Theta}^t, \boldsymbol{\Omega}^t, \boldsymbol{A}^t, \boldsymbol{F}^t)$: state variables,
      - $\boldsymbol{S}^t = (\boldsymbol{\Phi}^t, \boldsymbol{\Upsilon}^t)$: sensor variables.
    - Decomposition:
      - $P(\boldsymbol{\xi}^t \wedge \boldsymbol{\xi}^{t-\delta t} \wedge \boldsymbol{S}^t ... \wedge \boldsymbol{S}^0) =$
      - $P(\boldsymbol{S}^t|\boldsymbol{\xi}^t).P(\boldsymbol{\Omega}^t).P(\boldsymbol{A}^t).P(\boldsymbol{\Theta}^t|\boldsymbol{\Theta}^{t-\delta t} \wedge \boldsymbol{\Omega}^t)$
      - $.P(\boldsymbol{F}^t|\boldsymbol{\Theta}^t \wedge \boldsymbol{A}^t).P(\boldsymbol{\xi}^{t-\delta t} \wedge \boldsymbol{S}^{t-\delta t} ... \wedge \boldsymbol{S}^0)$
    - Parametric forms:
      - $P(\boldsymbol{S}^t|\boldsymbol{\xi}^t) = P(\boldsymbol{\Phi}^t|\boldsymbol{\Omega}^t).P(\boldsymbol{\Upsilon}^t|\boldsymbol{F}^t)$
        : sensor model, Gaussians and Dirac delta function;
      - $P(\boldsymbol{\Omega}^t),\ P(\boldsymbol{A}^t)$: a priori for state, Gaussians;
      - $P(\boldsymbol{\Theta}^t|\boldsymbol{\Theta}^{t-\delta t} \wedge \boldsymbol{\Omega}^t), P(\boldsymbol{F}^t|\boldsymbol{\Theta}^t \wedge \boldsymbol{A}^t)$
        : state dynamic model, Dirac delta functions;
      - $P(\boldsymbol{\xi}^{t-\delta t} \wedge \boldsymbol{S}^{t-\delta t} ... \wedge \boldsymbol{S}^0)$: previous iteration, distribution computed at last time step.
    - Identification:
      - Parameters of the Gaussians: σ_Φ, σ_Υ, σ_A and σ_Ω .
  - Question:
    - $P(\boldsymbol{\xi}^t|\boldsymbol{S}^t \wedge ... \wedge \boldsymbol{S}^0)$

Fig. 2. Bayesian program for processing of inertial data.

4) Bayesian inference: The *Bayesian Program* formalism, as first defined by Lebeltel [28], will be used throughout this text. The Bayesian program shown in fig. 2 will be used to compute the probability distribution for the current state given all the previous sensory inputs and initial known distribution.

To simplify notation, state variables are grouped in a vector $\boldsymbol{\xi}^t = (\boldsymbol{\Theta}^t, \boldsymbol{\Omega}^t, \boldsymbol{A}^t, \boldsymbol{F}^t)$ and sensor variables in a vector $\boldsymbol{S}^t = (\boldsymbol{\Phi}^t, \boldsymbol{\Upsilon}^t)$

At time t the Bayesian program computes the probability distribution of the current state $\boldsymbol{\xi}^t$ given all the previous sensory inputs the initial distribution $\boldsymbol{\xi}^t$. For the above Bayesian program the inference of current state is done by applying the conjunction and marginalization rule, applying a summation over state variables at the previous time step so that no decision is taken about these values, summarizing all the past in the answer to the estimation question in the previous time step, and can be formulated by:

$$\begin{aligned} P(\boldsymbol{\xi}^t|\boldsymbol{S}^t \wedge ... \wedge \boldsymbol{S}^0) = \quad & \tfrac{1}{K}.\textstyle\sum_{\xi^{t-\delta t}} P(\boldsymbol{S}^t|\boldsymbol{\xi}^t) \\ & .P(\boldsymbol{F}^t|\boldsymbol{\Theta}^t \wedge \boldsymbol{A}^t) \\ & .P(\boldsymbol{\Omega}^t).P(\boldsymbol{A}^t).P(\boldsymbol{\Theta}^t|\boldsymbol{\Theta}^{t-\delta t} \wedge \boldsymbol{\Omega}^t) \\ & .P(\boldsymbol{\xi}^{t-\delta t} \wedge \boldsymbol{S}^{t-\delta t} ... \wedge \boldsymbol{S}^0) \end{aligned}$$

where:

- K is a normalization constant;
- $P(\boldsymbol{S}^t|\boldsymbol{\xi}^t) = P(\boldsymbol{\Phi}^t|\boldsymbol{\Omega}^t).P(\boldsymbol{\Upsilon}^t|\boldsymbol{F}^t)$ is the sensor model, i.e., the probability distribution of sensor inputs given the sate. $P(\boldsymbol{\Phi}^t|\boldsymbol{\Omega}^t)$ and $P(\boldsymbol{\Upsilon}^t|\boldsymbol{F}^t)$ are Gaussians;
- $P(\boldsymbol{F}^t|\boldsymbol{\Theta}^t \wedge \boldsymbol{A}^t)$ a Dirac delta function, equal to 1 if and only if (2) is verified;
- $P(\boldsymbol{\Omega}^t)$, $P(\boldsymbol{A}^t)$ represent a priori knowledge about state variables, both Gaussians;
- $P(\boldsymbol{\Theta}^t|\boldsymbol{\Theta}^{t-\delta t} \wedge \boldsymbol{\Omega}^t)$ is the system dynamic model for state variable $\boldsymbol{\Theta}$, i.e., the probability distribution of rotation $\boldsymbol{\Theta}$ given

the previous rotation and current angular. $P(\mathbf{\Theta}^t|\mathbf{\Theta}^{t-\delta t} \wedge \mathbf{\Omega}^t)$ is a Dirac delta function, equal to 1 if and only if (1) is verified;
- $P(\boldsymbol{\xi}^{t-\delta t} \wedge \boldsymbol{S}^{t-\delta t} ... \wedge \boldsymbol{S}^0)$ is the probability distribution computed at last time step, i.e., from previous iteration of the Bayesian filter.

We can see also that the first-order Markov assumption is present in both the state dynamic model and sensor model: time dependence has a depth of one time step. The stationarity assumption is also implicit: models do not change with time. The filter iterates for each new time step, but the relationships between these variables remain the same for all time steps. This greatly reduces the complexity.

For the implementation the space of $\xi^{t-\delta t}$ that needs to be scanned has 3 dimensions: $\mathbf{\Theta}^{t-\delta t}$. For a given $\xi^{t-\delta t}$, the space of possible ξ^t has 3 dimensions, so the total search space has 6 dimensions. Resampling is applied, so that unlikely particles are deleted and likely ones are duplicated, in order to avoid having all particles drift towards improbable states. At each iteration a new set of N samples is drawn from the previous set of particles. Each particle of the previous set has a probability w^i to be chosen for each new particle. The weights in the new set to $1/N$.

C. Bayesian program for Image Stabilization and Gaze Control

Having the estimate for the current angular position and angular velocity of the system, we now need to implement a compensatory control so that the motors stabilize the image. The Bayesian program shown in fig. 3 implements the image stabilization for fixed gaze, using the inertial sensors data as input and the pan and tilt as actuators. This is an increment to the previous program that essentially adds a motor model to the Bayesian filter, with the motor model depending only on the current state and previous motor commands.

The actuator control is based on current angular position and velocity of the system. The pan and tilt units are controlled with combined commands for end position and velocity. The motors will move to the desired target position with the selected velocity and stop. The motor model takes this into account by having the current motor command depending on the current state and also on the previous motor commands. The remaining angular degree of freedom corresponds to rotating about the camera optical axis, and is done by digitally rotating the image after capture.

The following variables will be used: angular position composed of yaw, pitch and roll angles $\boldsymbol{\theta} = (y, p, r)$, $\mathbf{\Theta}$ is the corresponding rotation matrix. Therefore the state $\boldsymbol{\xi}$, as defined above, has all the information about $\boldsymbol{\theta}$; angular velocity, as defined above: $(\Omega_y, \Omega_p, \Omega_r)$; target pan&tilt: α_y, α_r; pan motor velocity and end position: $\mathcal{P}_\omega$,$\mathcal{P}_\theta$; tilt motor velocity and end position: $\mathcal{T}_\omega$, $\mathcal{T}_\theta$; and image rotation: $\mathcal{R}_\theta$.

Fig. 3 presents the Bayesian program of image stabilization for fixed gaze. To simplify notation, state variables are

Program
- Description
  - Specification
    - Relevant variables:
    - $\boldsymbol{\xi}^t = (\mathbf{\Theta}^t, \mathbf{\Omega}^t)$: state variables,
    - $\boldsymbol{M}^t = (\mathcal{P}_\omega, \mathcal{P}_\theta, \mathcal{T}_\omega, \mathcal{T}_\theta, \mathcal{R}_\theta)$: motor variables.
    - $\boldsymbol{S}$: sensor variables
    - $\boldsymbol{\alpha}$: target
    - Decomposition:
    - $P(\boldsymbol{\xi}^t \wedge \boldsymbol{M}^t|\boldsymbol{S}^t \wedge \boldsymbol{\alpha}^t) =$ $P(\boldsymbol{\xi}^t|\boldsymbol{S}^t).P(\boldsymbol{M}^t|\boldsymbol{\xi}^t \wedge \boldsymbol{\alpha}^t)$.
    - Parametric forms:
    - $P(\boldsymbol{\xi}^t|\boldsymbol{S}^t)$ direct from Bayesian program that processes inertial data ;
    - $P(\mathcal{P}_\omega|\boldsymbol{\xi}^t \wedge \boldsymbol{\alpha}^t)$, $P(\mathcal{T}_\omega|\boldsymbol{\xi}^t \wedge \boldsymbol{\alpha}^t)$, $P(\mathcal{P}_\theta|\boldsymbol{\xi}^t \wedge \boldsymbol{\alpha}^t)$, $P(\mathcal{T}_\theta|\boldsymbol{\xi}^t \wedge \boldsymbol{\alpha}^t), P(\mathcal{R}_\theta|\boldsymbol{\xi}^t \wedge \boldsymbol{\alpha}^t)$
    - : Gaussians with mean based on simple control.
  - Identification:
  - fixed control rules (or some learning of motor behaviour...)
- Question:
- $P(\boldsymbol{M}^t|\boldsymbol{S}^t \wedge \boldsymbol{\alpha}^t)$

Fig. 3. Bayesian program for motor control.

grouped in a vector $\boldsymbol{\xi}^t = (\mathbf{\Theta}^t, \mathbf{\Omega}^t)$ and motor variables in a vector $\boldsymbol{M}^t = (\mathcal{P}_\omega, \mathcal{P}_\theta, \mathcal{T}_\omega, \mathcal{T}_\theta, \mathcal{R}_\theta)$

A very simple deterministic proportional control was used for the motor control. The pan velocity counteracts the observed yaw angular velocity, limited by the maximum motor velocity, and the pan position angle has to take into account the desired target and the rotation between frames of reference. And likewise for tilt and pitch. Since the image rotation does not affect the IMU position, we simply rotate the image according to the estimated roll.

In order to format the stabilization control to a probabilistic framework, we considered Gaussian distributions about the mean given by the simplistic control. This will later be replaced by a better controller, to include different priors and even learning of the control, using the same Bayesian program.

The optical flow contribution will be added to enable gaze following of moving targets, and eventually as input for image cropping to have a stabilized output stream of images to overcome the dynamic limitations of the motors.

III. Results

A. Experimental Setup

The robotic gaze control has been implemented on an experimental setup using standard equipment, as shown in fig. 4. This setup, the Integrated Multimodal Perception and Experimental Platform (IMPEP), has been assembled to provide a common experimental platform for our work on multimodal perception computational models for visuo-auditory, visuo-haptic and visuo-inertial sensing.

For this work only the monocular camera, the inertial sensors and the pan and tilt motors will be used. The miniature inertial sensor used, Xsens MTi, provides digital

a)  b) 

Fig. 4. Integrated Multimodal Perception and Experimental Platform (IMPEP): a) stereo camera, monocular fast camera, inertial sensor, stereo microphones and laser range finder; b) mounted on pan&tilt and onto a camera tripod for manipulation.

output of 3D acceleration, 3D rate of turn (rate gyro) and 3D earth-magnetic field data. With this setup we will explore the use of the inertial sensor coupled directly to the camera as shown in fig. 4.

B. Results

For comparison with our Bayesian implementation, the Xsens IMU firmware MotionTracker was used to provide attitude estimation. The Motion Tracker implements a weighed filtering of the accelerometer, gyro and magnetic data to provide sensor angular position, including an adaptive filter used to correct for magnetic disturbances. The added data from the magnetic sensor enables the firmware estimation filter to provide a relative ground truth for our experimental work.

The above described Bayesian processing of inertial data was implemented in C with the ProBT library. Fig. 5 shows results for 1D single yaw rotation axis. The magnetic data enables the Xsens filter to outperform the Bayesian filter, as seen in the plot the probabilistic value accumulates some drift. Fig. 6 shows the result for pitch where gravity is taken into account by the Bayesian Filter to bound accumulated drift. Here a particle filter with just 200 particles was used, using $\sigma_A = 0.3m.s^{-2}$ and $\sigma_\Omega = 1rad.s^{-1}$. The system is initially calibrated [29], to have the rigid rotation between the inertial sensor and the camera, so that inertial measurements can be expressed in the egocentric vision system coordinates. The pan and tilt motors are aligned with the camera, providing camera yaw (pan) and pitch (tilt). The control system is running at about 2 Hz, limited by the pan&tilt hardware. The optical flow algorithm was implemented in C++ using

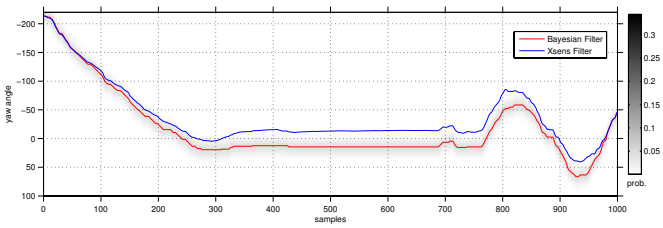

Fig. 5. Bayesian processing of inertial data, showing result for yaw angle at 100 Hz sample rate.

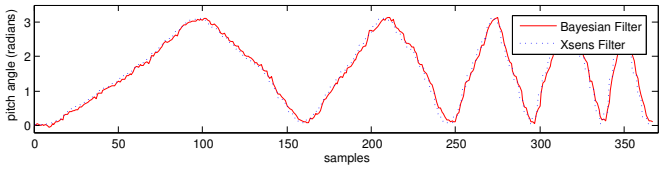

Fig. 6. Bayesian processing of inertial data, showing result for pitch where gravity is taken into account to bound accumulated drift.

the OpenCV library with Intel acceleration, obtaining $1.6fps$ on a standard PC.

Fig. 7 shows a set of results for image stabilization. Even though the signals have distinct sampling frequencies, the three plots denote the same time period. For every 100 samples from the IMU we have 30 images and slightly over 4 pan&tilt commands. We can see a strong correlation between the pan and yaw, and tilt and roll signals, since the pan&tilt is compensating the observed motion at a low sample rate.

The results show that the controller works but produces a noisy response due to the dynamic control limitations of the pan and tilt units. A new robotic vision head under construction will enable smooth motion control so that a better mechanical stabilization can be performed.

IV. CONCLUSIONS AND FUTURE WORK

We reported our work towards having a working robotic biomimetic implementation of image stabilization and gaze control based on using Bayesian inference. The Bayesian program was presented, and applied the robotic vision system. Current work is being done to optimize the optical flow code to have a full working real-time solution.

The IMU MotionTracker estimator, used for comparison in the results, factors in the magnetic data and provides a better estimate than the Bayesian processing of inertial data; however the Bayesian program will allow tuning to known priors, fusion with the visual input, and learning of motor behaviours.

Preliminary results show that the image stabilization yields satisfactory results, when subject to motions similar to what humans experience in normal conditions, but the current setup has clear limitations in dynamic response.

The optical flow contribution will enable gaze following of moving targets, and eventually as input for image cropping to have a stabilized output stream of images to overcome the dynamic limitations of the motors.

The construction of a a custom system with a better dynamic performance has been initiated. This binocular active vision system will allow stereo vergence with an adjustable baseline, with common head tilt and neck pan, mimicking the human degrees of freedom. Future work will address adding the magnetic data to our Bayesian implementation, providing a more robust attitude estimation.

V. ACKNOWLEDGMENTS

The authors gratefully acknowledge support from EC-contract number BACS FP6-IST-027140, and the help provided by José Prado, Hugo Faria and José Marinho on the technical implementation.

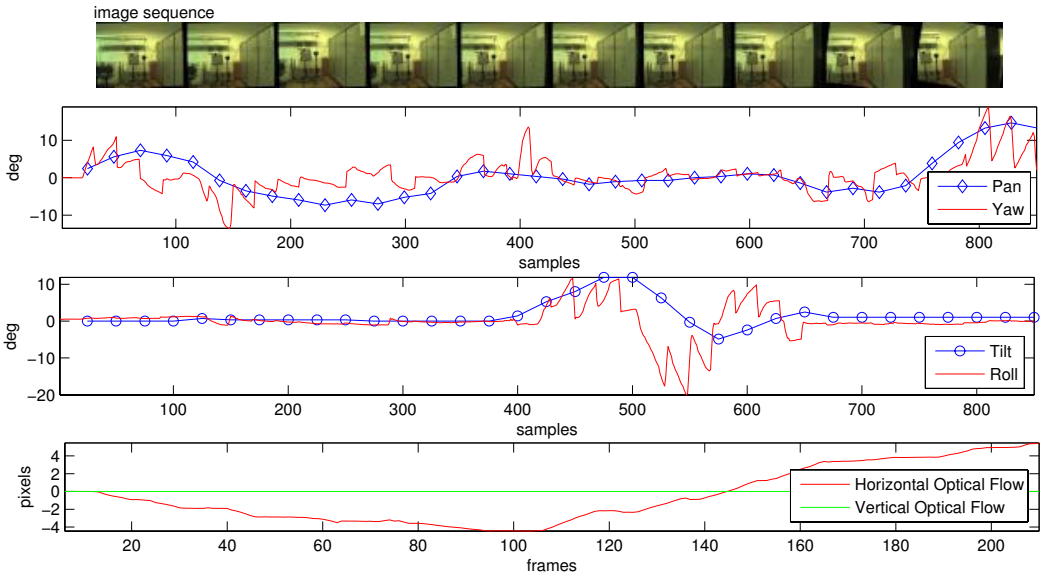

Fig. 7. Observed yaw and roll, pan and tilt motor control, and remaining observed optical flow.

REFERENCES

[1] H. Carpenter. *Movements of the Eyes*. London Pion Limited, 2nd edition, 1988. ISBN 0-85086-109-8.

[2] Alain Berthoz. *The Brain's Sense of Movement*. Havard University Press, 2000. ISBN: 0-674-80109-1.

[3] Peter Corke, Jorge Lobo, and Jorge Dias. An introduction to inertial and visual sensing. *The International Journal of Robotics Research (IJRR) Special Issue from the 2nd Workshop on Integration of Vision and Inertial Sensors.*, 26(6):519–535, June 2007.

[4] Jorge Lobo and Jorge Dias. Vision and inertial sensor cooperation using gravity as a vertical reference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1597–1608, December 2003.

[5] David C. Knill and Alexandre Pouget. The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12):712–719, December 2004.

[6] Yair Weiss, Eero P. Simoncelli, and Edward H. Adelson. Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6):598–604, 2002.

[7] Wilson S. Geisler and Daniel Kersten. Illusions, perception and Bayes. *Nature Neuroscience*, 5(6):508–510, June 2002.

[8] Francis Colas, Julien Diard, and Pierre Bessire. Common Bayesian Models For Common Cognitive Issues. Bacs deliverable 3.1: Review report of bayesian approaches to fusion multimodality, conflicts, ambiguities, hierarchies and loops, Bayesian Approach to Cognitive Systems (IP FP6-IST-027140), January, 11th 2007.

[9] Francesco Panerai and Giulio Sandini. Oculo-motor stabilization reflexes: integration of inertial and visual information. *Neural Networks*, 11(7-8):1191–1204, October 1998.

[10] Francesco Panerai, Giorgio Metta, and Giulio Sandini. Visuo-inertial stabilization in space-variant binocular systems. *Robotics and Autonomous Systems*, 30(1-2):195–214, January 2000.

[11] F. Panerai, G. Metta, and G. Sandini. Learning visual stabilization reflexes in robots with moving eyes. *Neurocomputing*, 48(1-4):323–337, October 2002.

[12] Stephane Viollet and Nicolas Franceschini. A high speed gaze control system based on the vestibulo-ocular reflex. *Robotics and Autonomous Systems*, 50(4):147–161, March 2005.

[13] Karl Pauwels, Markus Lappe, and Marc M. Van Hulle. Fixation as a mechanism for stabilization of short image sequences. *International Journal of Computer Vision*, V72(1):67–78, April 2007.

[14] Kikuro Fukushima. Corticovestibular interactions: anatomy, electrophysiology, and functional considerations. *Experimental Brain Research*, 117(1):1–16, October 1997.

[15] Susan Hurley. Perception and action: Alternative views. *Synthese*, 129(1):3–40, October 2001.

[16] Gilles Leone. The effect of gravity on human recognition of disoriented objects. *Brain Research Reviews*, 28(1-2):203–214, November 1998.

[17] D E Angelaki, M Q McHenry, J D Dickman, S D Newlands, and B J Hess. Computation of inertial motion: neural strategies to resolve ambiguous otolith information. *The Journal Of Neuroscience: The Official Journal Of The Society For Neuroscience*, 19(1):316–327, January 1999.

[18] Laurence R. Harris, Michael Jenkin, and Daniel C. Zikovitz. Visual and non-visual cues in the perception of linear self motion. *Experimental Brain Research*, 135(1):12–21, October 2000.

[19] Gilles Reymond, Jacques Droulez, and Andras Kemeny. Visuovestibular perception of self-motion modeled as a dynamic optimization process. *Biological Cybernetics*, 87(4):301–314, October 2002.

[20] Jonathan W. Kelly, Jack M. Loomis, and Andrew C. Beall. The importance of perceived relative motion in the control of posture. *Experimental Brain Research*, 161(3):285–292, March 2005.

[21] Jeremy Beer, Colin Blakemore, Fred Previc, and Mario Liotti. Areas of the human brain activated by ambient visual motion, indicating three kinds of self-movement. *Experimental Brain Research*, 143(1):78–88, March 2002.

[22] Fred H. Previc, Mario Liotti, Colin Blakemore, Jeremy Beer, and Peter Fox. Functional imaging of brain areas involved in the processing of coherent and incoherent wide field-of-view visual motion. *Experimental Brain Research*, 131(4):393–405, April 2000.

[23] Jean Laurens and Jacques Droulez. Bayesian processing of vestibular information. *Biological Cybernetics*, 96(4):389–404, April 2007.

[24] Peter Corke and M.C. Good. Dynamic effects in visual closed-loop systems. *IEEE Transactions on Robotics and Automation*, 12(5):671–683, October 1996.

[25] A. Pouget, P. Dayan, and R. Zemel. Information processing with population codes. *Nature Reviews Neuroscience*, 1:125–132, 2000. Review.

[26] John S. Zelek. Towards bayesian real-time optical flow. *Image and Vision Computing*, 22(12):1051–1069, October 2004.

[27] Ted Camus. Real-Time Quantized Optical Flow. In *IEEE Computer Architectures for Machine Perception*, September 18–20 1995.

[28] O. Lebeltel. *Programmation Baysienne des Robots*. PhD thesis, Institut National Polytechnique de Grenoble, Grenoble, France, September 1999.

[29] Jorge Lobo and Jorge Dias. Relative pose calibration between visual and inertial sensors. *The International Journal of Robotics Research (IJRR) Special Issue from the 2nd Workshop on Integration of Vision and Inertial Sensors.*, 26:561–577, 2007.