ELSEVIER

Neurocomputing 48 (2002) 323–337

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Learning visual stabilization reflexes in robots with moving eyes

F. Panerai[a,*], G. Metta[b], G. Sandini[b]

[a]Laboratoire de Physiologie de la Perception et de l'Action (LPPA), CNRS/Collège de France, 11 pl. M. Berthelot, 75005 Paris, France

[b]Laboratory for Integrated Advanced Robotics (LIRA), DIST, University of Genoa, Viale F. Causa 13, 16145 Genoa, Italy

Received 2 November 2000; accepted 6 June 2001

Abstract

This work addresses the problem of learning stabilization reflexes in robots with moving eyes. Most essential in achieving efficient visual stabilization is the exploitation/integration of different motion related sensory information. In our robot, self-motion is measured inertially with an artificial vestibular system and visually using optic flow algorithms. The first sensory system provides short latency measurements of rotations and translations of the robot's head, the second, a delayed estimate of the motion across the image plane. A self-tuning neural network learns to combine these two measurements and generates oculo-motor compensatory behaviors that stabilize the visual scene. We describe the network architecture and the learning scheme. The stabilization performance is evaluated quantitatively using direct measurements on the image plane. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Robotics; Artificial neural networks; Optic flow; Inertial information; Image stabilization; Oculo-motor reflexes; Sensori-motor maps

1. Introduction

Robot systems using vision as their primary sensorial modality often incur in the degradation of visual functions when self-movement or externally imposed disturbances occur. In humans and animals with moving eyes, degradation of visual

* Corresponding author. Tel.: +1-33-1-4427-1389; fax: +1-31-1-4427-1382.
E-mail address: francesco.panerai@college-de-france.fr (F. Panerai).

PII: S0925-2312(01)00645-2

functions does not occur owing to an efficient image stabilization mechanism. This mechanism operates generating compensatory eye movements that help in maintaining the image to be stable on the retina irrespective of the observer's motion. The brain circuitry responsible for this stabilization mechanism is the vestibulo-ocular reflex (VOR), further distinguishable in *rotational* VOR–RVOR and *translational* VOR–TVOR [14,23]. Two remarkable aspects of this circuitry are its presence and functionality from an early period of life [6] and secondly, its plasticity, i.e. its adaptive modification driven by any change which degrades visual stabilization performance. As a matter of fact, whatever the change is, an error signal is created which informs the brain that the VOR is not working properly. As a result, the system recalibrates itself [15]. This type of motor learning has been studied from different perspectives and models of the learning process and sites of learning have been proposed. The neural region that seems responsible for this kind of recalibration is the cerebellum [12]: the gain of the VOR reflex is nominally 1.0 and it is maintained closer to this value by a parametric-adaptive control system [25].

In robotics, image stabilization techniques dealing with multi sensory integration have received little attention so far. Since a decade, a growing number of studies have concentrated on active control of camera movements [1,11,2,24,4,18,22]. Most of these systems are very efficient [29,17] and perform well as static observers. On the other hand, robust performance when the robot agent plays a static observer in a "structured" environment, could be completely disrupted when operating in "disturbed, non-structured" conditions. How much of the robot's performance would be lost in the presence of external disturbances? To what extent the programmed functionalities would fail if the robot's vision system were onboard a vehicle moving on a rough terrain? What could be possibly envisaged to avoid degradation of visual performance? When external unpredictable disturbances intervene, most of the solutions based on "pure" visual processing, although engineered to a high level of performance, easily remain exposed to failure. On the contrary, approaches proposing the integration of inertial and visual sensory information have the advantage of being more robust [19,26]: visual functionalities, like motion detection or time to contact estimation are kept unaltered when external disturbances are present [19], tracking abilities do not degrade, saccades are more efficient, even in the presence of unpredictable movements [21]. When dealing with multi-sensory integration, instead of a pre-programmed context-dependent use of the different sensory inputs, it would be more general for the robot to learn such differences. Recently, Shibata and Schaal proposed an approach to image stabilization based on learned compensatory camera movements. In a monocular vision system, they have reproduced an accurate computational model of the biological VOR-OKR system. The authors show that the vision system can learn fast and optimally the oculo-motor compensatory behavior [26]. On the other hand, in a binocular robots—i.e. a robot controlling with vergence, the depth of the fixation plane—optimal stabilization would require the knowledge of the distance to the fixated object. In a previous work, dealing with geometric and computational issues, we have shown that optimal stabilization strategies (i.e. fine tuning of the oculomotor gains) can be implemented considering the geometry of the robot's binocular system (i.e. baseline, relative positions of eye and neck rotational axes) and the knowledge of actual gaze configuration (i.e. the gaze distance obtained through the vergence angle) [21]. In this work, we describe an alternative approach which consists in learning oculo-motor compensatory reflexes. The approach does not require an a priori computational model and simply relies on the availability of inertial and visual sensory information. A mixed neural architecture—growing neural gas (GNG) network and SoftMax function model—is used to build a sensory motor map which interpolates the two sensory signals (i.e. the inertial and visual motion information) and adjusts its parameters to generate optimal oculo-motor commands. The result of the learning is a sensori-motor map coding adaptively compensatory eye movements. The only assumption made at the beginning of the learning process is that the robot must be able to estimate the motion across the image plane and sense the head movement in terms of angular velocities (i.e. vertical and horizontal axes). The remainder of the paper is organized as follow: Section 2 introduces the notion of compensatory camera movements and shows how they could be used to reduce the visual motion; Section 3 describes the neural network model. Sections 4 and 5 elaborate on the learning scheme and its convergence behavior. Finally, Section 6 outlines the performance in image stabilization obtained with this approach.

2. Compensatory camera movements for image stabilization

In creatures with moving eyes, compensatory eye movements are the solutions adopted by biological evolution to deal with the visual consequences of head rotation and translation (Fig. 1). The sensory signals driving this fast oculo-motor

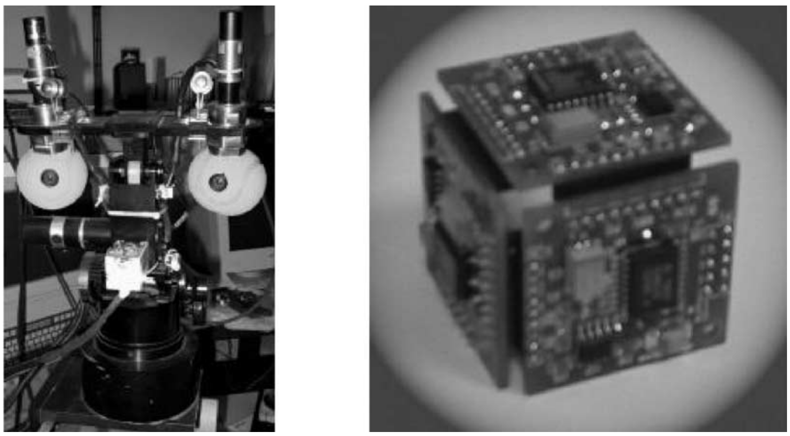

Fig. 1. The binocular robot head and the artificial vestibular system. Left: the robot head has 5 degrees of freedom, features space variant images and integrates an artificial vestibular system (i.e. small white cube centered below the robot's eyes). Right: the tri-axial version of the artificial vestibular system composed of three mutually perpendicular rotation modules (dimensions are $3 \times 3 \times 3$ cm^3).

response originate in the canals and otoliths organs of the vestibular system and are processed by a neural circuitry which implements a parametric-adaptive control system. For example, because of the 3D structure of the environment, images of objects at different distances require different amounts of compensatory eye movement [23,14]. Insights on these context-dependent issues are offered by the kinematic analysis of head translations, which shows that the eye movements that are required to eliminate from the images the translation consequences are inverse proportional to a particular depth plane [20]. For head rotations, though, dependence upon fixation distance is influential only at short distances and becomes negligible for longer distances. Moreover, during head rotations, the eccentric position of the eyes introduces a component of translation in the eye's movement, but the effect almost disappears for farther distances. This work concentrates on the problem of learning compensatory eye movements that eliminate the visual consequences of rotational head movements. In the next section, an optimal solution with respect to the robot's available sensory information is outlined.

2.1. Combining head movement and image motion information

The movement of a camera in 3D space produces an optic flow [28] that under simplifying hypotheses (i.e. limiting motion to one rotational component and ignoring second-order translational effects due to eye eccentricity) can be written as:

$$\frac{u_0}{f_x} = -W_y, \tag{1}$$

where W_y represents the rotation around a vertical axis and f_x the focal length of the camera. If we decompose the term W_y as the sum of the head velocity Ω_y, and the eye velocity with respect to head, w_y, we can rewrite Eq. (1) as:

$$\frac{u_0}{f_x} = -(\Omega_y + w_y) \tag{2}$$

Eq. (2) clearly establishes the relationship between the head rotational velocity, Ω_y and the eye compensatory movement w_y. Any misalignment between the two quantities will determine a residual motion u_0 on the image plane.

Relationship (2) can be inverted and used to learn the exact compensatory eye movements w_y required to stabilize the visual field (i.e. to zero the overall motion on the image plane). The robot estimates directly the head's rotational velocity Ω_y through its artificial vestibular system, and the optic flow u_0 through a visual processing module. The two measured quantities are independently fed into a neural network and the network will invert the previous relationship in trying to approximate the control surface representative of the optimal compensatory motor command w_y. The schematic diagram of Fig. 2 is representative of the information flow corresponding to this learning scheme. Until correct compensatory behavior is learned, a residual motion measured at the center of the image plane (i.e. u_0) represents, at the same time, a direct measure of the inadequacy of the compensatory motor response. To recover the horizontal and the vertical components u_0

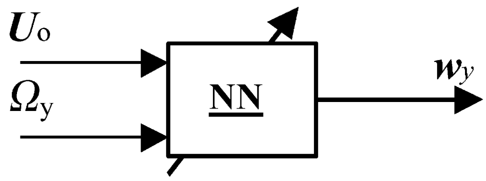

Fig. 2. Block diagram of the network inputs and output. The motion related sensory signals, the residual optic flow u_0 and the head's angular velocity Ω_y, are combined to produce a motor command, the compensatory eye movement w_y.

and v_0 (also called residual optic flow, ROF) of the image motion, a first order approximation or affine model of the optic flow is used [10]. From a computational point of view, the method proposed by Tunley and Young has been adopted [29]. In our robot, processing of visual information is implemented real-time (40 ms) on a double Pentium architecture.

3. The neural network model

The neural network used in our application is built on the GNG-Soft architecture [13]. It combines two existing models, namely, the growing neural gas (GNG) model [8] and SoftMax basis function networks [16], hence the name GNG-Soft. The GNG is an unsupervised network model, which learns topologies. A set of units connected by edges is distributed in the input space (a subset of $\Re^N$) using an incremental mechanism, which tends to minimize the mean distortion error: the distortion error is locally accumulated and a new unit is inserted near the unit with maximum error. At every learning step, a subset of units (the winner and its neighbors) is moved through the input space following a Hebbian learning rule. Among its properties, GNG is also able to adapt to the locally varying dimensionality depending indeed on the input data set. A detailed description of the algorithm can be found in [7]. A SoftMax function network consists of a single layer of *processing elements* (PEs) characterized by a receptive field centered on a preferred vector. Each unit's activation function U_i is a SoftMax function whose analytic expression is

$$U_i(x, c_i) = \frac{G(\|x - c_i\|)}{\sum_j G(\|x - c_j\|)}, \qquad G(x) = e^{-x^2/2\sigma^2} \tag{3}$$

where $G(\bullet)$ is a Gaussian function, $x \in \Re^N$ represents the input vector of the network, and c_i is the center of the activation function, that is, the input coordinate where the function U_i reaches its maximum. The index j extends to the number of Gaussian functions defined in the SoftMax model. The normalizing factor of U can be seen as a lateral inhibition mechanism [16]. The resulting hybrid architecture has the following advantages: first, the effectiveness in distributing the units within the multi-dimensional input space, typical of the GNG networks. Second, the "optimal" approximation and interpolation properties of SoftMax functions networks. Structurally, the network consists of a single layer of processing elements (PEs), each characterized by a receptive field-like structure. Benaim and Tomasini [3] proposed a Hebbian learning rule for the optimal placement of PEs:

$$\Delta c_i = \eta_1(x - c_i)U_i(x, c_i), \tag{4}$$

where $x \in \Re^N$ is an input pattern and η_1 the learning rate. Indeed, a SoftMax network can learn a smooth non-linear mapping $z = z(x)$. The reconstruction formula in this case is

$$z(x) \cong \sum_i v_i U(x, c_i), \tag{5}$$

where the parameters v_i are the weights of the output layer and $z \in \Re^M$. In particular, considering an approximation case, this formula can be interpreted as a minimum variance estimator [27]. The learning rule is

$$\Delta v_i = \eta_2(z - v_i)U_i(x, c_i). \tag{6}$$

This schema has been used to model cortical maps [5].

Finally, we devised a heuristic criterion in order to self-tune variances. At each learning step, the variance of the winning unit and its neighbors are updated using the following rule:

$$\sigma_i^2 = k\hat{d}_i^2, \tag{7}$$

where $\hat{d}_i^2$ is the mean squared distance from the unit i and its neighbors, and k is a positive constant. It is worth noting that the formula is largely similar to that suggested by Fritzke in [7]. A scaling factor (represented by k) has been added to guarantee a substantial overlap of the basis function tails. A description of this heuristics could be found elsewhere [13].

It is now evident as to how to combine the two self-organizing maps previously illustrated to obtain an incremental and plastic network with the best features of both models. The resulting model will be characterized by the effectiveness, typical of the GNG, in the distribution of the units within the n-dimensional input space and by the strength approximation and interpolation properties of SoftMax functions networks.

4. Learning principle and learning schemes

The learning scheme refers to a framework in which the robot has available sensory information about the instantaneous angular velocity of the head (i.e. Ω_y) and the first-order approximation of image motion (i.e. u_0). The input space of the network is therefore 2D, $x \in \Re^2$, $x = (\Omega_y, u_0)$. In order to adjust the network parameters, we have chosen as performance criterion, the residual motion measured on the image plane formalized in Section 2.1. Mathematically, the network has to minimize the following expression:

$$\min_{v_i, c_i} \left[- \left(\Omega_y + \sum_i v_i U_i(\xi, c_i) \right) \right], \tag{8}$$

where the sum over i represents the actual network output, the actual compensatory command (see Eq. (5)). We use the same formalism, but this time the input vector is named ξ. The term inside the square brackets of Eq. (8) is derived from the analytical expression of the optic flow obtained from the general flow equation for a moving observer under the constrained motion hypotheses (see Section 2.1); the addendum corresponding to the eye velocity has been substituted by the network output. In order to carry out the minimization, the weights of the output layer are modified according to the following Hebbian rule:

$$\Delta v_i = \eta_2 \left(\sum_i v_j U_j(\xi, c_j) + u_0 - v_i \right) U_i(\xi, c_i). \tag{9}$$

Note that the target output is shifted by the quantity u_0 from the current network output. In fact, whenever stabilization is perfect, $u_0 = 0$, no adjustment is necessary and in fact, $\Delta v_i \approx 0$. It is worth stressing that time, which is not explicitly indicated in Eq. (9), plays a fundamental role in this schema.

In fact, the residual optic flow used as input to the network is actually one time step before that used as a stabilization measure. That is, the measure of the network performance can be obtained only one step after the network has been used to generate a motion command. A delay line in Fig. 3 indicates this last point.

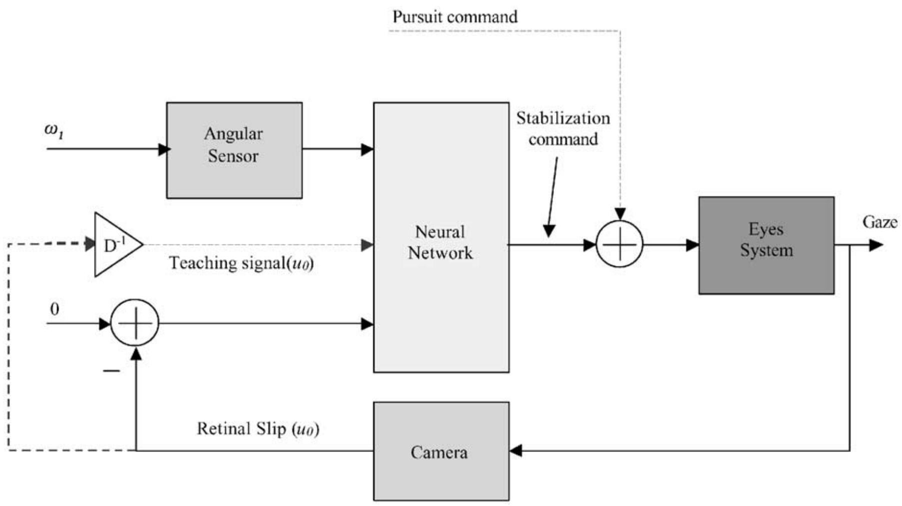

Fig. 3. Learning scheme. The inertial information, provided by the angular sensor, and the image motion (i.e. the retinal slip signal) are combined by the NN into the compensatory stabilization command. The teaching signal for the network is the residual image motion, which has to be minimized for stabilization to be effective.

5. Learning issues

5.1. Asymptotically stable behavior

In order to characterize the behavior of the proposed learning scheme, we analyzed at subsequent time intervals the distribution of the vectors (Ω_y, u_0) within the network input domain (Fig. 4). If the system learns the compensatory eye

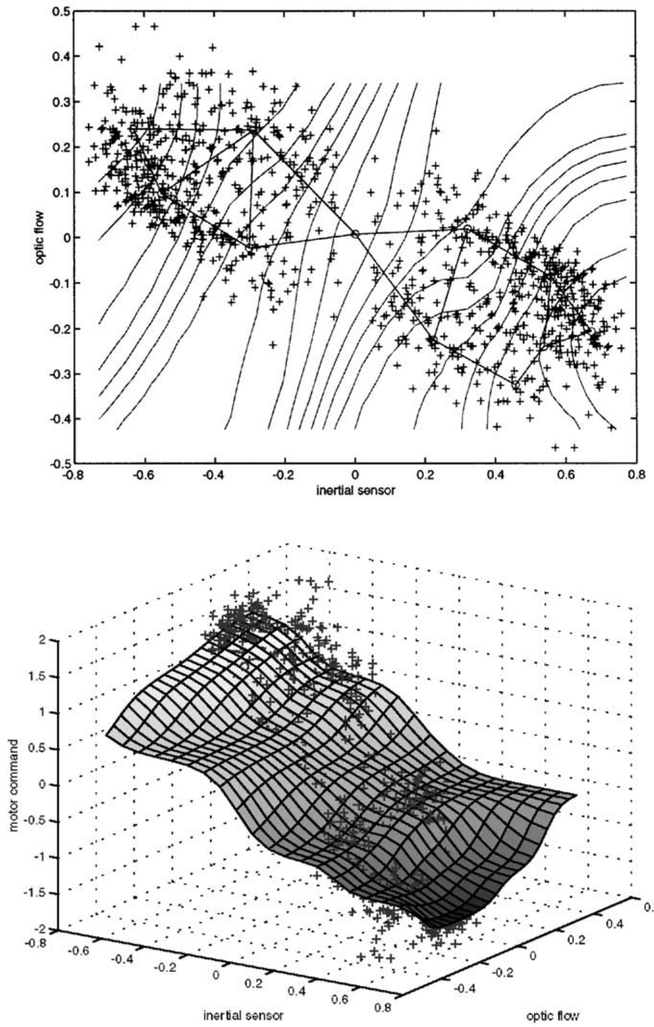

Fig. 4. The sensori-motor map of the stabilization reflex. Top: the distribution of the network units (light gray circles) in the input domain (x-axis inertial, y-axis optic flow). Bottom: the control surface representing the compensatory motor command (w_y).

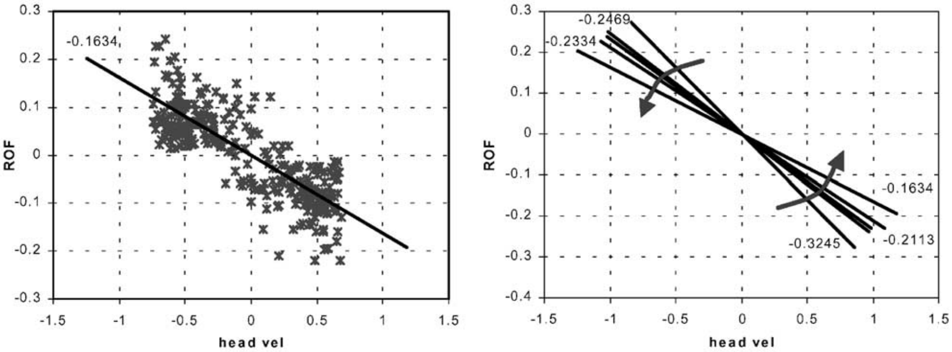

Fig. 5. Learning behavior of the neural controller in the input domain. Data are presented on normalized scales (normalization factors: 90°/s for the head velocity, 5 pixels/frame for the ROF). Left: a learning set, represented by the crosses, is characterized with a linear fit. The ratio between the inertial information (head velocity) and the residual optic flow (ROF) is modeled by the regression line coefficient ($a = -0.1634$). Right: Lines interpolating consecutive learning sets turn counterclockwise during the learning process.

movements correctly, then the residual motion (ROF) over learning trials will be asymptotically reduced. This is indeed the result of the process as shown in the two pictures of Fig. 5. Both pictures describe the two dimensional input domain of the neural controller. The horizontal axis codes the inertial measurement (i.e. the head angular velocity measured by the artificial vestibular system) and the vertical axis codes the residual image motion. In the left picture, the crosses identify data (about 300 samples) of one learning set. Their distribution is almost symmetrical with respect to the origin: a rotation of the head in one direction (e.g. positive) produces an opposite image drift (e.g. negative). We characterized the learning set using a linear regression. The slope of the interpolating line ($\alpha = -0.1634$) models the relationship between the amplitude of the head's rotational movement and the residual image motion. The behavior of the line's slope for consecutive trials is described in the right picture. The line's slope becomes smaller ($\alpha = -0.3245, -0.2469, -0.2334 \ldots -0.1634$) indicating that the learning scheme operates correctly over trials asymptotically reducing the image motion. The neural network was trained using motion stimuli of changing amplitude (a) and frequency (f) (e.g. $a = 0.3^{\circ}, 0.4^{\circ}, 0.5^{\circ}$/s and $f = 0.4$, 0.5, 0.6 Hz). In all such cases, a stable convergent behavior was achieved.

5.2. Learning curve

The learning scheme and learning behavior are further characterized in Figs. 6 and 7. In Fig. 6, a motion stimulus and the produced motor response are plotted on a 40 ms scale. The dark gray curve represents the rotational velocity of the head (*H_Vel*) as measured by the inertial sensor, while the light gray curve, the motor command (*E_Vel*) synthesized by the neural network during learning. The two

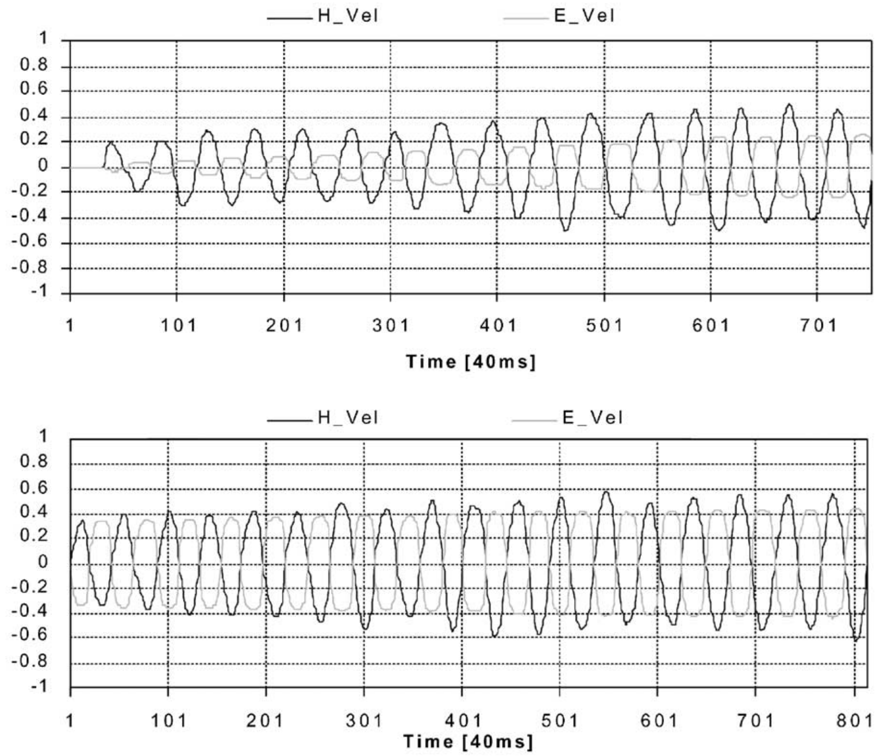

Fig. 6. Learning phases of the compensatory motor response. The two plots show how compensatory eye movements are acquired during learning. An external source of movement produces a repetitive rotational pattern (*H_Vel*) whose amplitude increases over time. The NN combines the two motion information and generates a compensatory command (*E_Vel*) to minimize the residual image motion. The two curves (*H_Vel* and *E_Vel*) are represented on a normalized scale (normalization factor is 90°/s). Time is expressed on a 40 ms scale. Top: an initial learning epoch in which the motor command (*H_Vel*) steadily increases over time: the compensatory oculo-motor response (*E_Vel*) is not yet effective. Bottom: an advanced learning epoch during which the compensatory motor response (*E_Vel*) is still growing, although at a slower rate.

pictures show the stabilization behavior during an initial and a more advanced phase of learning, defined, respectively, over 700 and 800 consecutive time stamps. Note that during the initial phase—top picture—the NN increases its output consistently, while in the bottom picture—the advanced phase—the NN still increases the motor response but at a lower rate.

Fig. 7 left, plots the learning curve for the initial phase, representing the residual angular velocity error defined as $Err = (1 - E\_Vel/H\_Vel)$. The dotted profile on the right picture, is the residual image motion (ROF) concurrently measured. Note that by the end of the 20th learning cycle, the NN is already capable of generating a compensatory motor command, which reduces the image motion induced by the motion stimuli of about 80%. Note also that at the same time, the ROF

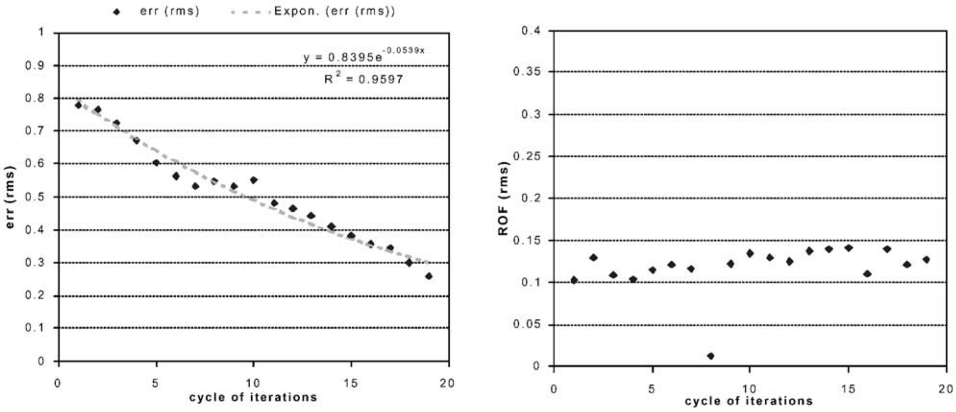

Fig. 7. Learning curve and ROF. A numerical evaluation of how well the system is learning. Left: The dotted profile represents the error (RMS value) measured during learning throughout 20 cycles of iteration. Error is computed as the ratio between the RMS values of compensatory motor command and the amplitude of the external motion stimuli. The dashed light gray curve is the exponential function $a\exp(-bx)$ which interpolates the dotted profile ($a = 0.83$, $b = 0.053$, $R^2 = 0.95$). Right: the ROF (RMS value) during learning remains bounded to 1 pixels/frame (0.15×5 pixels/frame).

is maintained below the threshold of 0.5 pixels/frame during the whole learning process.

6. Stabilization performance

To evaluate the impact of external disturbances on the visual functionalities of the system a performance criterion based on visual measurements is of paramount importance. Residual optic flow (ROF) is therefore the performance index selected to evaluate quantitatively the efficiency of the stabilization mechanism. Two different external motion stimuli are used. In all the performance measurements, the sensory motor map acquired as described in Section 4 was used to generate compensatory eye movements. The two pictures in Fig. 8 represent on a normalized scale (i.e. normalization factors are, respectively, 90°/s for the head angular velocity and 5 pixels/frame for the residual visual motion) the measurements corresponding to the two different motion stimuli (stimuli frequency and amplitude are, respectively, 0.3 Hz, 18°/s amplitude and 0.6 Hz, 81°/s amplitude). The amount of ROF is evaluated in correspondence to the maximum peak velocity of each stimulus (diamonds markers on the light gray curve, triangular markers on the dark gray curve). In the top picture, the peak head velocity measured by the inertial sensor reaches values of about 18°/s and the average peak amplitude of the residual motion is about 0.5 pixel/frame. In the bottom picture though, the amplitude and the frequency of the external motion change substantially (i.e. frequency roughly doubles and amplitudes increase four times approximately). Note that even in this second case, the amount of image motion still remains bounded to 1 pixel/frame.

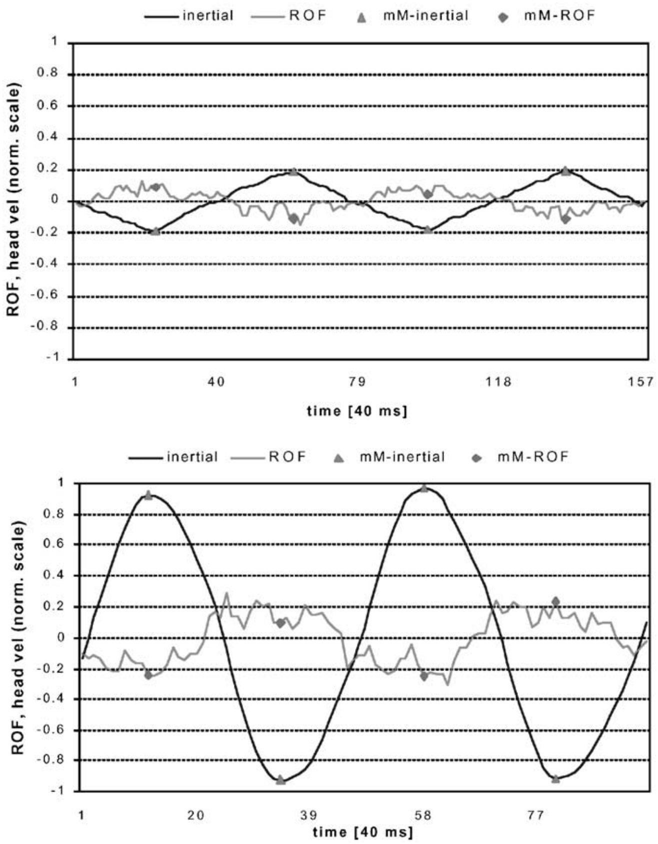

Fig. 8. Image stabilization performance measured through the ROF. The plots represent on a normalized scale the residual optical flow (ROF) measured during the occurrence of the motion stimuli (normalization factor is 5 pixels/frame). Top: with a motion stimulus of about 18°/s amplitude, the learned sensori-motor map produces a compensatory response which keeps the ROF bounded to less than 0.5 pixel/frame (0.1×5 pixels/frame). Bottom: with an external motion of about 90°/s amplitude, the learned compensatory eye movements keep the ROF bounded to 1 pixel/frame (0.2×5 pixels/frame).

In other words, the network performs efficiently in both cases, responding to a five-fold increase of the amplitude of the motion stimulus (and a doubling of the frequency) with only a small increase in the residual image motion, i.e. from 0.5 pixel/frame to 1 pixel/frame.

7. Conclusion

A framework to learn oculo-motor control for image stabilization purposes has been described. The key point in achieving satisfactory stabilization performance over a wide range of external or imposed motion/disturbances is the use of two different types of motion related sensory information. Inertial and visual cues related to self-motion are the sensory signals required to stabilize the visual world efficiently. A method to learn to combine this sensory information is proposed. It is based on the auto-organizing properties of two classes of artificial neural networks, namely the GNG model and SoftMax function networks. An unsupervised learning scheme enables the robots to build a sensory motor map transforming self-motion signals into oculo-motor compensatory commands. The learning scheme is efficient in adapting the network parameters and becomes effective after a short training period. The stabilization performance obtained with this approach has been evaluated numerically and directly on the image plane: the residual visual motion remains bounded to 1 pixel/frame for externally imposed disturbances of different dynamic characteristics. Although in the current paper we described the results obtained for image stabilization in the horizontal plane, the approach could be easily extended to generate compensatory reflexes for both vertical and horizontal axes.

Concerning critical issues about the current approach, we have observed that the generation of the compensatory oculo-motor command suffers from the constant delays present in the visual (40 ms) and inertial (30 ms) loops. The main effect of these delays results in a phase lag of the compensatory command generated by the network with respect to the externally imposed disturbance (respectively, *E_Vel* and *H_Vel* in Fig. 6). Both delays are constant, therefore improvements can be obtained using simple linear prediction or more sophisticated techniques such as Kalman filtering (see for example [9]). An alternative solution, not investigated in the present paper, would be to learn the delays using a tapped-delay-line approach (see for example [26]). Moreover, although the current implementation achieves satisfactory stabilization performance, there are at least two possible sources of errors which influence the generation of the compensatory motor command. The primary source of error derives from the saturation of the measurement of the zero-order component of the optic flow (i.e. u_0). This effect is due to the implementation issues of the optic flow algorithm. For the current optic flow algorithm, the saturation effect was previously assessed both in terms of range and accuracy [19]. A secondary source of error may derive from the simplified model of the optic flow which neglects the second-order translational effects due to the eye eccentricity. These effects are considered relevant especially for short fixation distances. We are considering the development of solutions which deals with such an effect, by extending the present learning approach to account for different fixation distances.

Acknowledgements

This work was supported by Italian Space Agency (ASI) and the EU Esprit Project NARVAL.

References

[1] J. Aloimonos, I. Weiss, A. Bandopadhay, Active vision, Internat. J. Comput. Vision 1 (1988) 335–356.

[2] D.H. Ballard, C.M. Brown, Principles of animate vision, Comput. Vision Graph. Image Process. 56 (1992) 3–21.

[3] M. Benaim, L. Tomasini, Competitive and Self-organinzing algorithms based on the minimization of an information criterion, in: T. Kohonen, K. Makisara, O. Simula, J. Kangas (Eds.), Artificial Neural Networks, North-Holland Elsevier Science Publisher, Amsterdam, 1991, pp. 391–396.

[4] C. Capurro, F. Panerai, G. Sandini, Dynamic vergence using log-polar Images, Internat. J. Comput. Vision 24 (1997) 79–94.

[5] R. Durbin, G. Mitchison, A dimension reduction framework for understanding cortical maps, Nature 343 (1990) 644–647.

[6] D.V. Finocchio, K.L. Preston, A.F. Fuchs, Infant eye movements: quantification of the vestibulo-ocular reflex and visual-vestibular interactions, Vision Res. 31 (1991) 1717–1730.

[7] B. Fritzke, Fast learning with incremental RBF Networks, Neural Process. Lett. 1 (1994) 2–5.

[8] B. Fritzke, A growing neural gas network learns topologies, in: G. Tesauro, D.S. Touretzky, T.K. Leen (Eds.), Advances in Neural Information Processing Systems, MIT Press, Cambridge MA, USA, 1995.

[9] A. Kiruluta, M. Eizenman, S. Pasupathy, Predictive head movement tracking using a Kalman filter, IEEE Trans. Syst. Man, Cybernet.-Part B: Cybernet. 27 (1997) 326–331.

[10] J. Koenderink, J. Van Doorn, Affine structure from motion, J. Opt. Soc. Amer. 8 (1991) 377–385.

[11] E.P. Krotkov, Active computer vision by cooperative focus and stereo, Springer, Berlin, 1989.

[12] S.G. Lisberger, Physiological basis for motor learning in the vestibulo-ocular reflex, Otolaryngol. Head, Neck Surg. 119 (1998) 43–48.

[13] G. Metta, Babyrobot: A Study an Sensory-motor Development, LIRA Lab, DIST, Ph.D. Thesis, University of Genova, 1999.

[14] F.A. Miles, C. Busettini, U. Schwarz, Ocular responses to linear motion, in: H. Shimazu, Y. Shinoda (Eds.), Vestibular and Brain Stem Control of Eye, Head and Body Movements, Springer-Verlag/Japan Scientific Societies Press, Tokyo, 1992, pp. 379–395.

[15] F.A. Miles, J.H. Fuller, Adaptive plasticity in the vestibulo-ocular responses of the rhesus monkey, Brain Res. 80 (1974) 512–516.

[16] P. Morasso, V. Sanguineti, Self-organizing body schema for motor planning, J. Motor Behav. 27 (1995) 52–66.

[17] D.W. Murray, K.J. Bradshaw, P.F. McLauchlan, I.D. Reid, P.M. Sharkey, Driving saccade to pursuit using image motion, Internat. J. Comput. Vision 16 (1995) 205–228.

[18] P. Nordlund, T. Uhlin, Closing the loop: pursuing a moving object by a moving observer. Sixth International Conference on Computer Analysis of Images and Patterns, 1995.

[19] F. Panerai, Sandini, Oculo-motor stabilization reflexes: integration of inertial and visual information, Neural Networks 11 (1998) 1191–1204.

[20] F. Panerai, G. Metta, G. Sandini, Visuo-inertial stabilization in space-variant binocular systems, Robotics Autonomous Syst. 30 (2000) 195–214.

[21] S. Rougeaux, Real-time Active Vision for Versatile Interaction, Ph.D. Thesis, Université d'Evry, Courcouronne, France, 1999.

[22] U. Schwarz, F.A. Miles, Ocular responses to translation and their dependence on viewing distance. I. Motion of the observer, J. Neurophysiol. 66 (1991) 851–864.

[23] P.M. Sharkey, D.W. Murray, S. Vandevelde, I.D. Leid, P.F. McLauchlan, A modular head/eye platform for real-time reactive vision, Mechatronics 3 (1993) 517–535.

[24] M. Shelhamer, D.A. Robinson, H.S. Tan, Context-specific gain switching in the human vestibuloocular reflex, Ann. New York Acad. Sci. 656 (1992) 889–891.

[25] T. Shibata, S. Schaal, Biomimetic gaze stabilization, Word Scientific, Singapore, 1999.

[26] D.F. Specht, Probabilistic Neural Networks, Neural Networks (1990) 109–118.

[27] V. Sundareswaran, Egomotion from global flow field, IEEE Workshop on Visual Motion. Princeton, NJ, USA, 1991.
[28] H. Tunley, D. Young, First order optical flow from log-polar sampled images. Proceedings of the Third European Conference on Computer Vision (ECCV), Stockholm, 1994.
[29] T. Uhlin, P. Nordlund, A. Maki, J. Ekhlund, Towards an active visual observer, Sixth International Conference On Computer Vision, 1995, pp. 679–686.

Francesco Panerai graduated in Electronic Engineering in 1993 at University of Genova. He obtained his Ph.D. in Robotics and Computer Vision at LIRA Lab. His thesis developed the idea of integrating visual and inertial information for the stabilization of gaze in a binocular robot. From 1998, he is a research associate at LPPA (CNRS/Collège de France), where he joined a team working on psychophysics of vision and movement. His current interests include multi-modal sensory integration for the perception of 3D space, the perception of distances and velocities in driving simulators.

Giorgio Metta holds an MS degree and Ph.D. in Electronic Engineering from the University of Genova, Italy. He worked on a humanoid robot from a biologically motivated perspective, with the ultimate goal of learning how to model biological agents by building complex artificial systems. Since 2001, he is a Postdoctoral Associate at the MIT, AI-Lab. His current interests are man-machine interaction, imitation learning and the development of gesture/language in robotics.

Giulio Sandini teaches the course of Natural and Artificial Intelligent Systems for students of the Electronics and Computer Science curriculum. He currently coordinates the activity of researchers at the Laboratory of Integrated Advanced Robotics (LIRA Lab) where research related to robotics and computational neuroscience is carried out. The ongoing projects include the control of a binocular head using space-variant, anthropomorphic sensors and the guidance of mobile robots on the basis of visual information. Giulio Sandini has been a member of program committees of international conferences and chairman and co-chairman of international conferences and workshops.