Machine Learning, 31, 187–200 (1998)

Emergence and Categorization of Coordinated Visual Behavior Through Embodied Interaction

LUC BERTHOUZE AND YASUO KUNIYOSHI {berthouz,kuniyosh}@etl.go.jp
Electrotechnical Laboratory(ETL), AIST, MITI
Intelligent Systems Division
Umezono 1-1-4, Tsukuba, Ibaraki 305, Japan

Editors: Henry Hexmoor and Maja Matarić

Abstract. This paper discusses the emergence of sensorimotor coordination for ESCHeR, a 4DOF redundant foveated robot-head, by interaction with its environment. A feedback-error-learning(FEL)-based distributed control provides the system with explorative abilities with reflexes constraining the learning space. A Kohonen network, trained at run-time, categorizes the sensorimotor patterns obtained over ESCHeR's interaction with its environment, enables the reinforcement of frequently executed actions, thus stabilizing the learning activity over time. We explain how the development of ESCHeR's visual abilities (namely gaze fixation and saccadic motion), from a context-free reflex-based control process to a context-dependent, pattern-based sensorimotor coordination can be related to the Piagetian 'stage theory'.

Keywords: foveated active vision, oculomotor control, feedback-error-learning, emergent coordination, sensorimotor memory

1. Introduction

Human babies are born with a rich set of innate behavioral abilities (Thelen & Smith (1994); Meltzoff & Moore (1989)). Immediately after birth, they are engaged in complex sensorimotor interaction with the world through their bodies. Through the interaction, babies acquire novel coordination skills which are used for further exploration of the world, introducing new classes of interaction. And the cycle goes on. The latter view of the developmental process was extensively examined by Jean Piaget (1962), who took it to an extreme and crystallized it as his 'stage theory'. We do not support Piaget's uniform and rigid stage theory. And we do support that innate abilities are much more than what Piaget assumed. However, we believe that some of his core ideas discussed below are still valid, useful, and important for adaptive autonomous agents, even in the light of rich innate abilities.

1.1. Hypotheses on Embodied Sensorimotor Development

This paper focuses on two important learning principles involved in the above bootstrapping developmental process. First, emergence of coordinated behavior by interacting with the world through its body, guided by innate reflexes. Second, emergent categorization of acquired coordinated behavior, which will result in categorical (selective and prototypical, or entrainment) responses to novel sensorimotor patterns. Together, they constitute one bootstrapping step for acquiring novel behavior by unsupervised, autonomous exploration of embodied interaction with the world. Constraints imposed by innate reflexes and the body structure are quite important, because on one hand, the world is too complex for *tabula rasa* learning, and on the other hand, it is unrealistic to provide complete supervisory signals for a continuously moving robot exhibiting unpredictable variation of motoric patterns.

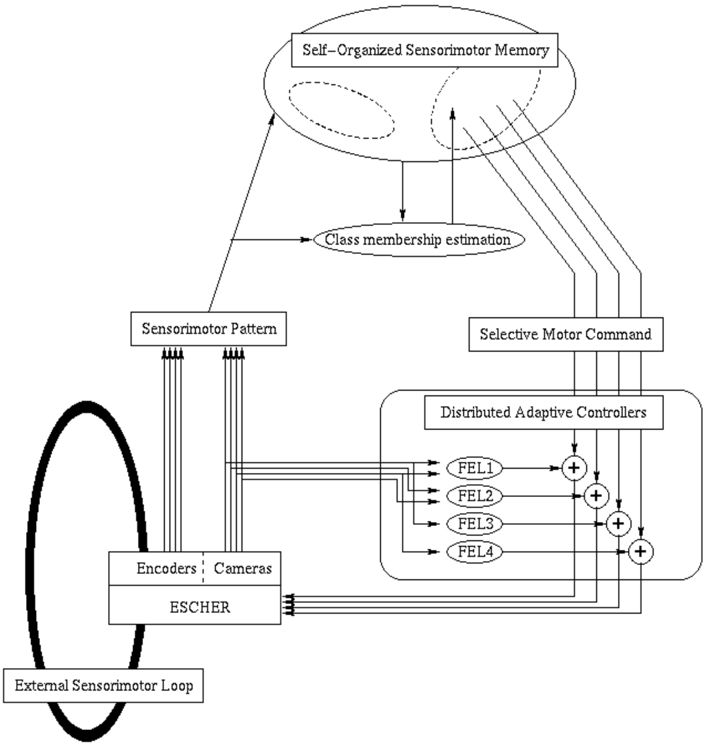
Self–Organized Sensorimotor Memory
Class membership estimation
Sensorimotor Pattern
Selective Motor Command
Distributed Adaptive Controllers
FEL1
FEL2
FEL3
FEL4
Encoders
Cameras
ESCHER
External Sensorimotor Loop

Figure 1. Overview of the architecture proposed in this paper. The distributed FEL (Feedback-Error-Learning (see text)) controllers learn the sensorimotor loop between ESCHeR and the environment by emerging a coordination. The self-organized sensorimotor memory generate selective responses to a new sensorimotor pattern based on previous experiences.

Our hypotheses examined in this paper are two-fold:

(1) Coordinated behavior can emerge from a set of distributed adaptive controllers which are independently interacting with the environment through the shared body. Because the body imposes consistent constraints on the interaction, the coordination can be achieved even when there is no explicit internal connections among the controllers. Each adaptive controller has crude innate knowledge (reflex) which adds a useful constraint for stabilizing and guiding the unsupervised learning. Section 3 presents our model and experimental results using distributed FEL(Feedback Error Learning) modules.

(2) Autonomous categorization of thus acquired coordinated behaviors is possible to some degree, because the body imposes certain consistent structure on the interaction dynamics. And the categorized representation can be used to generate selective responses to a given

sensorimotor pattern. This can be used to start the next step of exploration in an effectively constrained manner. This last step is interesting especially because we use a vision system as a test-bed. Given a visual motion pattern, such as waving a hand, the system first reacts by acquired tracking behavior, then categorizes self motion, which in turn can generate an ‘interpreted’ motor response to the initial stimulus. This is a form of behavioral imitation. And the system has the potential ability to learn how to imitate a novel behavior, which is an important component of real world intelligence (Kuniyoshi 1994). Section 4 presents our method and experimental results using a Kohonen SOM (self organizing map).

1.2. Focus of the Paper

As an experimental platform for examining the above hypotheses and corresponding methods, we chose a binocular robotic head system which is capable of active gaze control (described in section 2). This system is ideal for exploring the above ideas because it is a compact but complete system for embodied sensorimotor interaction; it reacts to visual stimuli by gaze movements, which affects the visual input, thus closing the sensorimotor loop. Moreover, it has nonlinearity and redundancy which introduces interesting challenges for sensorimotor coordination which are not present in simple wheeled mobile robots.

Our emphasis in the present paper is on investigating the global hypotheses and principles discussed above: The component learning algorithms (FEL and SOM) used in this paper are not novel, because our focus is not on individual learning algorithms, but rather on how they interact with each other and with the world through the body, leading to an emergence of novel performance. Although our experimental results show some useful performance achievements and some inspirations related to biological systems, neither optimal performance and effectiveness in the engineering sense, nor faithful replication of a biological system are our main concern.

2. ESCHeR: *E*tl *S*tereo *C*ompact *He*ad for *R*obot Vision

As the platform for our experiments, we adopted a robotic vision head, called ESCHeR (Kuniyoshi, Kita, Rougeaux, & Suehiro, 1995). ESCHeR is a 4 DOF (degrees of freedom) binocular active vision mechanism: as shown in Fig. 2, it has two CCD cameras which rotate independently (“vergence”) in a common horizontal plane which can be tilted, and the whole platform can rotate around the vertical axis (“pan”). All joints are driven by DC servo motors equipped with rotary encoders. The mechanism partially mimics the eye mobility of human vision system. ESCHeR is provided with a high performance gaze mobility (close to humans), which is sufficient for tracking a moving object (“smooth pursuit”) or to quickly change the focus of attention (“saccade”).

The lowest level control (such as achieving a commanded velocity) of the motors is done at 500Hz cycles by a servo controller, a dual Transputer (IMS T805) system, which communicates motion commands and proprioceptive data with higher level controllers via 20Mbps serial communication channels. Real time image processing is implemented using a DataCube MaxVideo system, a pipeline architecture which does fast preprocessing, and a KIWIVision system, a distributed transputer system (9CPU’s) which does post processing and communication with the servo controller. They are connected via MaxBus image data bus, and controlled by a MVME167 Motorola 68040 based CPU board running LynxOS, a real time UNIX. The learning programs presented in this paper were implemented in part (which requires real time operation) on the servo controller CPU's, and the rest (which does not need strict real time operation) on a workstation connected with the above system via Ethernet.

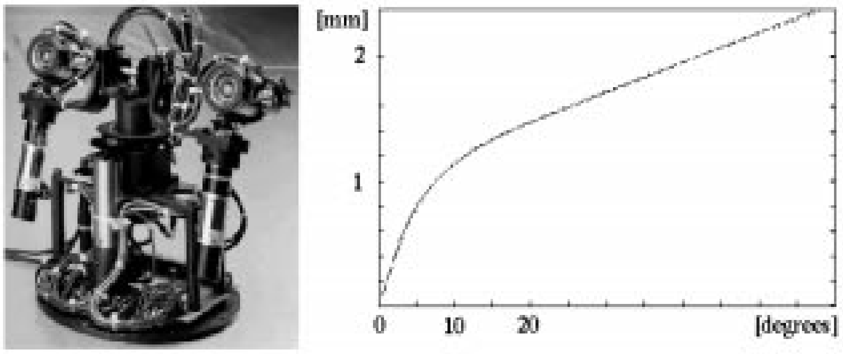

Figure 2. ESCHeR: a 4DOF robot-head (left) with foveated vision. (right) ESCHeR's nonlinear projection curve.

The most significant part of ESCHeR is its "foveated wide angle" lenses (Kuniyoshi, Kita, Sugimoto, Nakamura, & Suehiro, 1995). They simulate human visual system's compromise between a wide but low resolution field of view for peripheral detection and a tiny high-resolution *fovea*[1] for precise observation. Our lens seamlessly combines the above two extreme characteristics in a single special optics which implements a nonlinear projection curve (Figure 2(right)). It has 120° field of view and the maximum magnification of 7.7 in the fovea versus the periphery.

The projection curve is a combination of three parts: (1) the fovea (with incident angles between 0° and 2.5° from optical axis) adopts a standard projection, (2) the periphery (from 20° to 60°) adopts a spherical projection, and (3) the intermediate range (2.5° to 20°) adopts a log-of-spherical projection. The log component can be combined with a polar transformation, which has many useful characteristics (Sandini & Tagliaso, 1980). It enhances the lock-on effect in stereo fixation, simplifies the analysis of optical flow and introduces image invariance to rotation and scaling (a powerful characteristics for 2D identification).

3. Learning Gaze Control

3.1. *Justification*

Control of active vision systems (see Rougeaux & Kuniyoshi (1997) for the state of the art) typically relies on two main visuomotor processes: gaze fixation (also called tracking or visual pursuit) and saccadic motion. In gaze fixation, the gaze platform is controlled so as to keep a given environmental cue in the center of the field of view, independently of its possible motion. A saccade can be defined as a fast conjugate change of eye position between fixations.

In terms of control theory, the visuomotor control on ESCHeR is relevant to the control of a 4DOF nonlinear redundant manipulator: it has a nonlinear image-to-joint Jacobian because of the optics' nonlinearities and it has redundancy in horizontal rotations of vergence and pan.

Past approaches to tracking/saccading control fall into two categories: (1) preprogramming (hard-coded solutions) such as (Murray, Bradshaw, McLauchlan, Reid and Sharkey, 1995; Nordlund & Uhlin, 1995; Brown, 1988; Rougeaux, Kita, Kuniyoshi, Sakane & Chavand, 1994) and (2) *tabula rasa* approach as in Smagt & Krose (1991). Both approaches exhibit drawbacks. Explicit programming is difficult, lacks flexibility to deal with novel behaviors and does not easily scale up (e.g., addition of new degrees of freedom). Intermediate approaches such as combining a limited amount of *a priori* knowledge with learning have not been extensively explored in robot learning, especially in the context of nonlinear redundant systems.

3.2. Feedback-error-learning

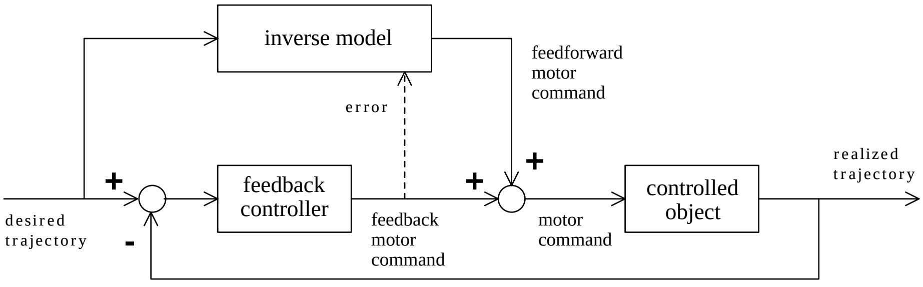

Figure 3. Feedback error learning: reproduced from (Kawato & Gomi, 1992).

The main obstacle to the implementation of a real-time adaptive approach is the determination of an *error-signal*. Described in Figure 3, feedback-error-learning, first introduced by Kawato, Furukawa & Suzuki (1987), provides us with an elegant solution. *Innate knowledge* about the controlled object is introduced under the form of a conventional feedback controller (CFC). Modeling a linear approximation of the inverse-model of the controlled-object, its output is taken as the error signal and is propagated (dotted arrow on the figure) onto a neural network inverse model (NNIM). The motor command estimated by NNIM is summed to the output of CFC so that NNIM does not mimic CFC but rather acquires the nonlinearities of the plant by minimizing the output of CFC. NNIM's synaptic modification follows the general Widrow-Hoff rule – approximation of a gradient descent –, assuming that CFC's motor output plays the role of error-signal.

The interest of such an approach stems from the following points:

- The desired plant output is used both for control and training, hence allowing an on-line learning. Instead, in other approaches such as *direct inverse modeling* (as described by

(Jordan, 1992)), the controller has to be trained off-line because the input of the controller is the actual plant output, and not the desired plant output.

• Although the training data that the system receives – pairs of actual plant inputs and desired plant outputs – are not samples of the inverse dynamics of the plant, the system nonetheless converges to an inverse model of the plant because of the error-correcting properties of the feedback controller. Unlike a predictive controller, the feedback controller does not require the implementation of an explicit plant model but rather a qualitative knowledge of the plant. Besides, it is shown that the performance of an error-correcting controller is generally rather insensitive to the exact value of the gain that is chosen (Jordan, 1992).

• Under assumptions that are fully detailed by (Gomi & Kawato, 1993), the global convergence of the scheme can be proven, using German's theorem and Lyapunov's second method. The two main conditions are: (1) a guaranteed asymptotic convergence of the learning phase (role of the gain in the linear controller) and (2) a very small and positive-definite learning rate.

3.3. Control of the Gaze Platform

With the architecture depicted in Figure 1, *an* image-to-joint transfer function is learned within the closed-loop between the visual information provided by the cameras and the motor information given by the encoders. In our experiments (such as described below), two types of visuomotor information have been considered, with similar results:

• Control in position by respect to the visual displacement of an environmental cue in the field of view. In such case, a given environmental cue (a pin-point light) is characterized by the position of its center of mass. At each frame, its displacement is computed and fed to each of the controllers described below. The output of the controllers is a command in position [2].

• Control in velocity by respect to the optical flow resulting from the combined motion of both the environmental cue and ESCHeR's gaze platform.

As in (Rougeaux & Kuniyoshi, 1997), we have a real-time optical flow computation based on Lucas & Kanade (1981)'s method. The extraction of the target's velocity is made through the following steps: (1) The background is assumed static, therefore, any peripheral flow can only be generated by ESCHeR's motion. Its component is extracted and subtracted from the general flow. (2) If a moving cue is in the field of view, the center of mass of the corresponding peak is computed and its flow extracted. The components of this flow is fed in the controllers which return a command in velocity.

Unlike most engineering-oriented approaches (Murray et al., 1995; Nordlund & Uhlin, 1995; Brown, 1988; Coombs, 1992; Rougeaux & Kuniyoshi, 1997) or VOR [3], each of ESCHeR's four joints (pan, tilt, left and right vergence) is *independently* controlled by a feedback-error-learning controller whose conventional feedback controller (CFC) is a crude (low gain) linear controller. The tuning of these controllers (the value of their gain ranging from 1×10^{-4} and $3 \times 10^{-4} rad/s$ in our implementation) results from a tradeoff between (1) the convergence of each controller (the linear controller guides the learning of the adaptive component of each controller) and (2) the overall stability of the architecture (higher gain values result in higher sensitivity to delays).

The adaptive component of each controller is a classical three-layer feed-forward artificial neural network. Both input and output layers are linear and dedicated to normalization. The hidden layer uses nonlinear activation functions (arctangent sigmoids). Synaptic modification is achieved using a back-propagation with momentum (Rumelhart, Hinton & Williams, 1986). While any Newton-like method (such as in MLP (multi-layer perceptron), CMAC (Cerebellar Model Articulator Controller) or RBF (Radial Basis Function)) would be acceptable, the use of back-propagation made a real-time implementation possible (reduced computational load). The weight-decay technique (Chow & Teeter, 1994) is used to prevent from a degenerescence of the synaptic weights, which usually follows a continuous adaptation.

3.4. Active fixation and Saccadic motion

The visual stimulus is given to ESCHeR in the form of a manually moved pin-point light. Throughout the experiment, one eye (the left one) is kept under the control of the conventional feedback controller (CFC) of its controller. This eye is used as a base of reference for the eye (right) which is under adaptive control. A new frame is acquired every $33ms$. After about $20s$ of continual visual stimulus, the neural component (NNFC) of each controller becomes dominant. Figure 4 shows the variance of the visual error (the quadratic sum of each coordinate) for each eye. It can be observed that the variance is significantly reduced for the eye under adaptive control. When evaluated in term of angular error (the conversion from pixels to angles being made using a spatial resolution curve derived from the projection curve), the residual error never exceeds 2°. This residue is, in part, the result of processing delays: $50ms$ between the image acquisition and the realization of the corresponding motor command. It could be partially compensated for by using predictive filtering.

In the following experiment, bright/moving objects are (spatially and temporally) randomly presented to ESCHeR in the periphery of its field of view (with at least 20° incidence angle, or 140 pixels from the center). Figure 5 shows the resulting visual error for each eye, the left one being kept under linear control. For the eye under adaptive control (right eye), the average acquisition time is 5 times shorter than for the reference eye: a 20° saccade takes only 3 frames or $100ms$, $60ms$ being the average time for a human over the same saccade.

The results above confirm that ESCHeR has learned to track *and* saccade within a single control framework. This result partially follows from ESCHeR's optical design: while the adaptive component of the controllers enable the generalization of the tracking skill to objects appearing in the periphery, the generalization is facilitated by the foveated vision through which objects presented in the periphery *appear* to be closer than they would in a conventional camera.

3.5. Emergence of a coordination

Applying unconstrained motor learning techniques to nonlinear redundant manipulator is generally avoided because the risk of getting trapped in stable but inconsistent minima increases with the dimension of the learning space. In our experiments, the introduction of each degree of freedom is delayed. This is justified by the biological notion of "freezing of

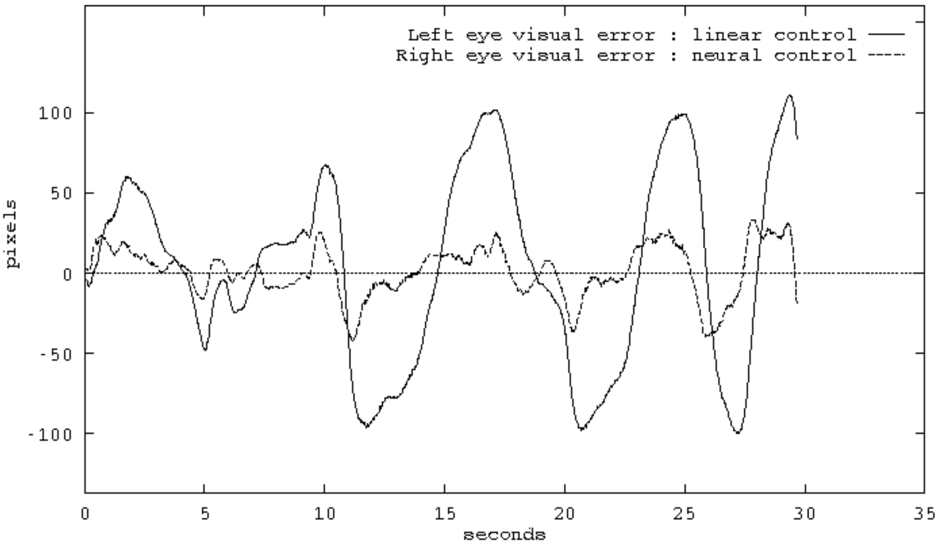

Figure 4. Comparison between visual error in left and right cameras during a tracking experiment: the left eye is kept under (fixed gain) linear control; the right eye has learned (neural component of its controller is dominant).

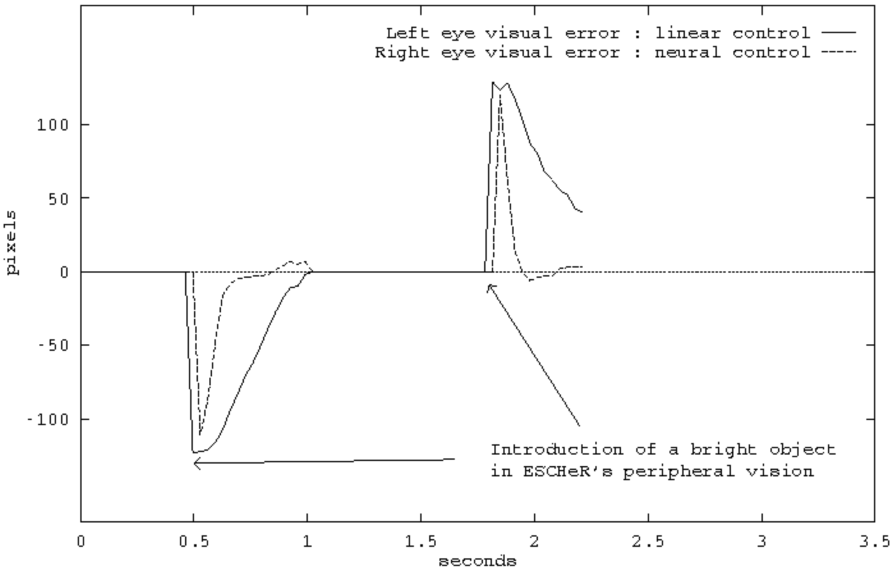

Figure 5. Bright objects that appear randomly in the field of view (at 0.5 and 2 seconds), attract ESCHeR's attention. The right eye (under adaptive control) achieves a faster acquisition of the novel objects than the eye kept under linear control.

degrees of freedom”. The visual stimulus is given in the form a swinging pendulum. At first, vergences only are controlled, until a stable control is learned (qualitative estimation). The control of the redundant joint (pan joint) is then introduced. As shown in Figure 6, a period of instability, with high-frequency vibrations, follows after which, a stable coordination (conjugate vergence and pan motion) emerges.

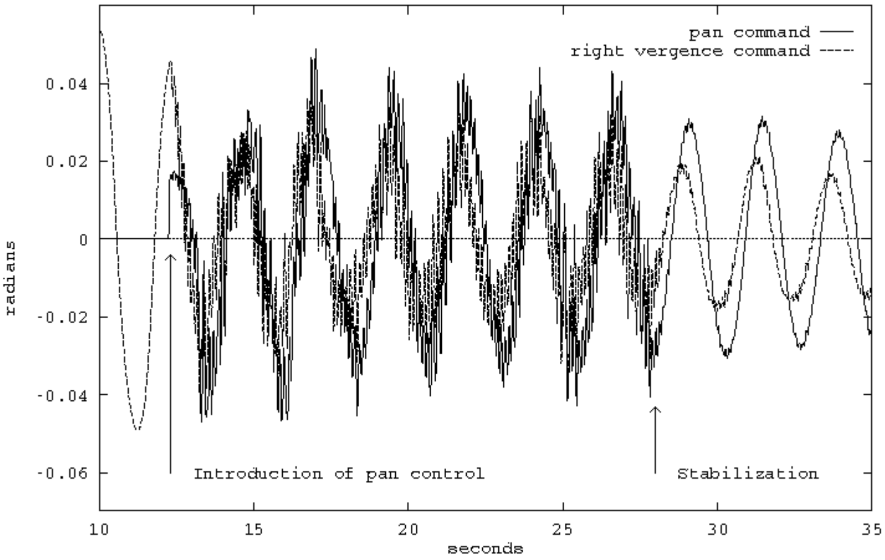

Figure 6. Learning the redundancies: concurrence of pan and vergence control introduces redundancies. High-frequency vibrations are observed until a VOR-like control law is learned.

Four factors can be accounted for this result:

• the flexibility of a distributed adaptive architecture: instead of being rigidly and explicitly coupled, controllers are implicitly coupled to each other, through the sensorimotor loop. The adaptive component that enables them to learn the nonlinearities of their own sensorimotor loop also allow them to compensate for the interferences generated by the motor commands of the other controllers.

• the feedback-error-learning approach: whereas the control space is generally randomly sampled until an acceptable plant output is found, in feedback-error-learning, the conventional feedback controller serves to guide the system to the correct region of the control space, *by making an essential use of the error between the desired plant output and the actual plant output* (Jordan & Rosenbaum, 1989).

• the delay of introduction of the redundant joint: it reduces the complexity of learning for each joint and allow the stabilization of the adaptive parameters of each controller.

• the periodicity of the stimulus: it contributes to the stability of the learning by way of reinforcement.

With the emerging coordination, ESCHeR can successfully track a moving target, as described in Figure 7. Additionally, although no calibration, neither optical nor mechanical

Pan control	Vergence control	Std dev (pixels)
Not controlled	linear	80
	neural	22
Controlled	linear	50
	neural	13

Figure 7. Standard deviation of the visual error during tracking experiments with various setups.

were performed, it is possible to extract the image-to-motor Jacobian from ESCHeR's behavior.

3.6. Conclusion

The experimental results reported in this section demonstrate that, given a set of base behaviors and a crude control system amenable to feedback, ESCHeR has developed fine oculo-motor control in interaction with its environment. It allowed it to *learn* how to track moving objects with relatively good precision and engage in saccading at speed comparable to humans. The continuous adaptation, however, leads to a short-term memory system. Discontinuities of the visual stimulus lead to a *reset* of the synaptic weights and thus of the whole system. In next section, we enable ESCHeR (1) to recognize behaviors that it already knows and (2) to reproduce them. The precedence of this reproduction over an unnecessary adaptation leads to the reinforcement of past actions and contribute to the long-term stability of the learned sensorimotor coordination.

4. Sensorimotor Similarity and Long-term Memory

4.1. Introduction

The problem of the representation is a critical issue. A state-space determined based on human intuitions is not necessarily appropriate for the robot. Instead, the state-space should be self-constructed through a statistical analysis of one's sensorimotor patterns. In our experiments, *self-organizing maps* (SOM) are used to perform a unsupervised categorization (identification) of ESCHeR's sensorimotor patterns over its (active) interaction with the environment. Self-organizing maps (SOM) were preferred to Learning Vector Quantization (LVQ) so that the repertoire of possible visual behaviors would not be predefined.

A self-organization map represents the result of an algorithm of vector quantization that places a number of reference vectors (or codebook vectors) (on which data sets will eventually be approximated) into a high-dimensional data space in an ordered fashion. When local-order relations are defined between the reference vectors, the relative values of the latter are made to depend on each other as if their neighboring values would like along an *elastic surface*. A mapping can be defined that projects this high-dimensional space

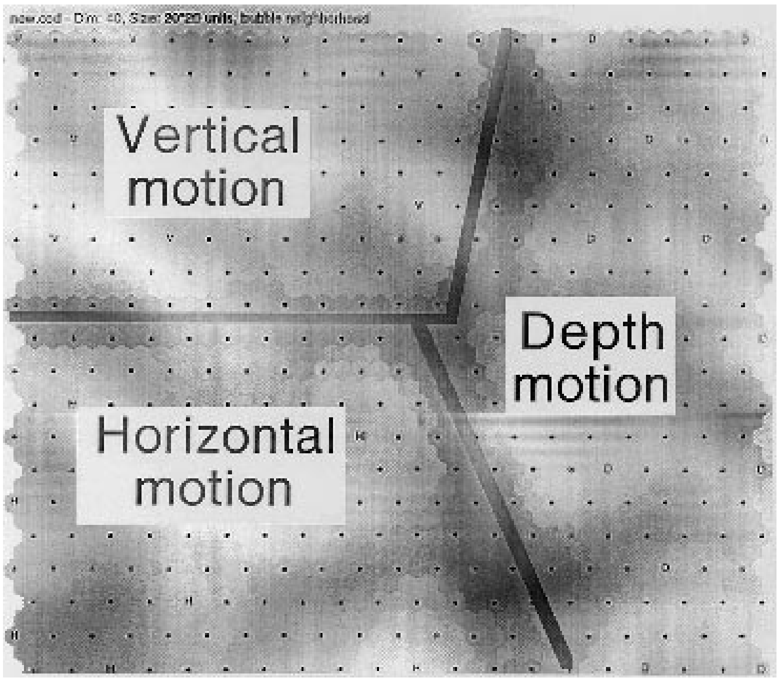

Figure 8. Two-dimensional representation of inter-class distances after the self-organization process has converged. This representation is generated using Kraaijveld, Mao & Jain (1992)'s technique. High intensities of grey indicate the boundary between clusters (high inter-distance). Four emergent sensorimotor classes have been highlighted.

onto a two-dimensional lattice of points. Such lattice is very suitable to the analysis of processes states (e.g., finding clusters within a set of sensorimotor patterns or identifying an unknown data vector with a cluster) because it allows the visualization of metric ordering relations between input samples (of high-dimensional order). In term of implementation, the mapping is obtained as an asymptotic state in a learning process (Kohonen, Hynninen, Kangas & Laaksonen, 1995).

4.2. Generation of reference vectors

The generation of reference vector is made as follows: ESCHeR is repetitively given visual stimuli (circular motion, horizontal or vertical waving gesture for example), different in amplitude and length. Each gesture is preceded by a blank period so that its starting and ending points can be easily determined. Over one gesture, the sequences of both the motor commands of each joint (pan, tilt, left and right vergences) and the components of the corresponding optical flow are recorded at each frame [4]. Once the gesture is finished, each sequence is normalized and discreetized in a fixed number (10) of samples (so as to remove both the temporal and the spatial dimension). Finally, all the sequences over one period are butted against each other to create a fixed-length sensorimotor pattern which is fed into the self-organizing process.

4.3. Recognition and reproduction

The self-organization process leads to a sensorimotor categorization such as shown in Figure 8. Four classes can be distinguished (the labels resulting from an off-line analysis). Three of them identify motions in orthogonal directions of the 3D space (horizontal, vertical and in-depth motions). The fourth class is uncertain at that stage. With the integration of new sensorimotor patterns, it may be reinforced as part of the class identified as horizontal motion or emerge as a new class, intermediate between motion in the horizontal plan and motion in depth [5].

In the online process however, such analysis is not carried out but the identification of any new sensorimotor pattern extracted over ESCHeR's environmental interaction with the emerged categorization is made by computing its *quantization error*. The probability of class membership is determined using a *k-nearest neighbors*-type method. When the probability of membership is sufficiently high (arbitrarily fixed level in our experiments), a generic motor command is computed on the average of each sensorimotor patterns belonging to that class. Once fed to the joint controllers, this motor command will result in a selective and prototypical response to the given sensorimotor pattern.

Conceptually, this is a higher-order replication of the *reflex vs adaptivity* paradigm on which each FEL controller is based. It guides and stabilizes the overall learning process over time.

5. Conclusion

Presenting their view about Active and Exploratory Perception, Bajcsy & Campos (1992) write: *We have attempted to conceptualize the perceptual process of an organism that has the top-level task of surviving in an unknown environment. During this conceptualization process, four necessary ingredients have emerged for either artificial or biological organisms. First, the sensory apparatus and processing of the organism must be active and flexible. Second, the organism must have exploratory capabilities. Third, the organisms must be selective in its data acquisition process. Fourth, the organism must be able to learn.*

The above points ground the work presented in this paper. With ESCHeR, the robot is endowed with an active sensory apparatus, whose optical characteristics make it flexible enough to handle close fixation as well as peripheral detection. The multiple independent feedback-error-learning based controllers guarantee explorative capabilities of the control, both at the individual level (with the neural network component of each controller) as well as at the collective level since the joint coordination is driven by a dynamic environmental interaction and each controller's plasticity. Through the self-organization by Kohonen network of its sensorimotor patterns, ESCHeR's progressively refines the categorization of its self interacting with the environment and can ultimately become selective in its data acquisition process through his newly acquired sensorimotor memory. With reflexes supervising the learning (through feedback controllers embedded in each controller), a stable sensorimotor coordination emerges (within a continuous environmental interaction) in which ESCHeR's visual abilities (gaze fixation and saccadic motion) reach a good level of performance.

This emergence of sensorimotor coordination leads us to suggest this approach as a first milestone towards *learning by imitation* in which an agent learns by observation of, and active participation in, the environment.

Notes

1. The fovea is the rod-less part of the human retina. It covers about 0.5° (in diameter) of the average 160° of entire field of view.
2. The frequency of the motor commands being higher than the frame rate achieved by ESCHeR, an intermediate control in velocity is achieved so that the requested position is reached at the next frame acquisition.
3. The animal *vestibulo-ocular-reflex* (VOR) stabilizes the retinal images during a neck movement by compensatory eye movements. For a target at infinite distance, it is achieved by causing the motion of the eyes to be equal and opposite to the motion of the head. It can be seen as a transformation from head velocity to eye velocity (Jordan 1990). When the target is not at infinite distance, a real-time adaptation is achieved for which Kawato & Gomi (1992) propose a computational model based on feedback-error-learning.
4. Such a high frequency was however not necessary. Consistent results were eventually achieved when recording every 3 frames ($100ms$).
5. In that experiment precisely, this class was generated by horizontal circular motions.

References

Bajcsy, R., & Campos, M. (1992). Active and Exploratory Perception. *CVGIP: Image Understanding*, 56(1), pp. 31–40.

Brown, C. (1988). *The Rochester Robot* (Technical Report TR-257). University of Rochester, Rochester, USA.

Chow, M., & Teeter, J. (1994). An analysis of Weight Decay as a Methodology of Reducing Three-Layer Feedforward Artificial Neural Networks for Classification Problems". *IEEE International Conference on Neural Networks*, pp. 600–605.

Coombs, D. (1992). *Real-Time Gaze Holding in Binocular Robot Vision.* Doctoral dissertation (also available as TR-415), Department of Computer Science, University of Rochester, Rochester, USA.

Gomi, H., & Kawato, M. (1993). Neural Network Control for a Closed-loop System Using Feedback-Error-Learning. *Neural Networks*, 6, pp. 933–946.

Jordan, M.I. (1990). Motor Learning and the Degree of Freedom Problem". In Jeannerod, M. (ed) *Attention and Performance*, vol. XIII, pp. 796–836.

Jordan, M.I. (1992). *Computational Aspects of Motor Control and Motor Learning* (Technical Report TR-9206). Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences, USA.

Jordan, M.I., & Rosenbaum, D.A. (1989). Action. In M.I. Posner (Ed.), *Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Kawato, M., Furukawa, K., & Suzuki, R. (1987). A Hierarchical Neural-Network Model for Control and Learning of Voluntary Movement. *Biological Cybernetics*, 57, pp. 169–185.

Kawato, M., & Gomi, H. (1992). A Computational Model of Four Regions of the Cerebellum Based on Feedback-error Learning. *Biological Cybernetics*, 68, pp. 95–103.

Kohonen, T., Hynninen, J., Kangas, J. & Laaksonen, J. (1995). *SOM-PAK: the Self-Organizing Map Program Package* (Technical Report April 7). Helsinki University of Technology, Laboratory of Computer and Information Science, Rakentajanaukio 2 C, SF-02150 Espoo, Finland.

Kraaijveld, M.A., Mao, J., & Jain, A.K. (1992). A Non-linear Projection Method Based on Kohonen's Topology Preserving Maps. *International Conference on Pattern Recognition*, Los Alamitos, CA, pp. 41–45.

Kuniyoshi, Y., Kita, N., Sugimoto, K., Nakamura, S., & Suehiro, T. (1995). A Foveated Wide Angle Lens for Active Vision. *IEEE International Conference on Robotics and Automation*, Japan, pp. 2982–2985.

Kuniyoshi, Y., Kita, N., Rougeaux, S., & Suehiro, T. (1995). Active Stereo Vision System with Foveated Wide Angle Lenses. In S.Z.Li, D.P. Mital, E.K. Teoh, H. Wang (eds) *Recent Developments in Computer Vision*, Lecture Notes in Computer Science 1035, Springer-Verlag, pp. 191–200.

Kuniyoshi, Y. (1994). *The Science of Imitation — Towards Physically and Socially Grounded Intelligence —*. RWC Technical Report, TR-94001.

Lucas, B., & Kanade, T. (1981). An Iterative Image Registration Technique With an Application to Stereo Vision. *Proc. DARPA Image Understanding Workshop*, pp. 121–130.

Meltzoff, A.N., & Moore, M.K. (1989). Imitation in Newborn Infants: Exploring the Range of Gestures Imitated and the Underlying Mechanisms. *Developmental Psychology*, vol. 25, no. 6, pp. 954–962.

Murray, D.W., Bradshaw, K. J., McLauchlan, P.F., Reid, I.D., & Sharkey, P.M. (1995). Driving Saccade to Pursuit using Image Motion. *International Journal of Computer Vision*.

Nordlund P., & Uhlin, T. (1995). *Closing the Loop: Detection and Pursuit of a Moving Object by a Moving Observer* (Technical Report CVAP-175-95-7-173). Computational Vision and Active Perception Laboratory, Royal Institute of Technology, S-100 44 Stockholm, Sweden.

Piaget, J. (1962). *Play, Dreams and Imitation in Childhood*. New York: W. W. Norton.

Rougeaux, S., & Kuniyoshi, Y. (1997). Velocity and Disparity Cues for Robust Real-Time Binocular Tracking. *IEEE Proc. Computer Vision and Pattern Recognition*, Puerto-Rico, pp. 1–6.

Rougeaux, S., Kita, N., Kuniyoshi, S., Sakane, S., & Chavand, F. (1994). Binocular Tracking Based on Virtual Horopters. *IEEE Proc. International Conference on Intelligent Robots and Systems*, Munich, Germany, pp. 2052–2057.

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning Representations by Back-propagating Errors. *Nature*, vol. 323, pp. 533–536.

Sandini, G., & Tagliaso, V. (1980). An Anthropomorphic Retina-like Structure for Scene Analysis. *Computer Graphics and Image Processing*, 14(3), pp. 365–372.

Smagt, P., & Krose, B.J.A. (1991). A Real-time Learning Neural Robot Controller. *International Conference on Neural Networks*, Espoo, Finland, pp. 351–356.

Thelen, E., & Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*, Cambridge, Mass.: MIT Press, Bradford Books.

Received September 1, 1997
Accepted December 30, 1997
Final Manuscript February 1, 1998