

An Approach Toward Self-Organization of Artificial Visual Sensorimotor Structures

Jonas Ruesch and Ricardo Ferreira and Alexandre Bernardino

Abstract Living organisms exhibit a strong mutual coupling between physical structure and behavior. For visual sensorimotor systems, this interrelationship is strongly reflected by the topological organization of a visual sensor and how the sensor is moved with respect to the organism's environment. Here we present an approach which addresses simultaneously and in a unified manner i) the organization of visual sensor topologies according to given sensor-environment interaction patterns, and ii) the formation of motor movement fields adapted to specific sensor topologies. We propose that for the development of well-adapted visual sensorimotor structures, the perceptual system should optimize available resources to accurately perceive an observed phenomena, and at the same time, should co-develop sensory and motor layers such that the relationship between past and future stimuli is simplified on average. In a mathematical formulation, we implement this request as an optimization problem where the variables are the sensor topology, the layout of the motor space, and a prediction mechanism establishing a temporal relationship. We demonstrate that the same formulation is applicable for spatial self-organization of both, visual receptive fields and motor movement fields. The results demonstrate how the proposed principles can be used to develop sensory and motor systems with favorable mutual interdependencies.

Key words: sensorimotor coupling, morphological adaptation, self-organization

1 Introduction

Visual perception is often considered a one-way process which passes a recorded stimulus along a sensory pathway at the end of which a conclusion is reached regard-

J. Ruesch, R. Ferreira, A. Bernardino
Instituto Superior Técnico, Institute for Systems and Robotics, Computer and Robot Vision Laboratory, Av. Rovisco Pais 1, 1049-001 Lisbon, e-mail: {jruesch,ricardo,alex}@isr.ist.utl.pt.

ing the observed scene. However, by simply observing an animal relying on visual perception in its natural environment, it becomes immediately clear that the animal’s motor apparatus is permanently engaged in supporting perception by orienting and relocating the visual sensory organs. Motor and sensory systems are working in a very close relationship where not only visual input affects future actions, but motor actions also actively contribute to the process of perception by “shaping” the sequence of recorded stimuli. Thus, in living organisms, the process of visual perception does not merely consist of visual stimuli being analyzed along the sensory pathway, but must be considered as a closed sensorimotor loop in which the animal’s body plays an important role [13]. In developmental psychology, this point of view has most prominently been advocated by Gibson [5]. But also more recently, O’Regan and Noë argue that visual percepts are acquired through training and execution of so-called sensorimotor skills [11]. In their view, a visual percept is “created” through exploration of sensorimotor contingencies during interaction with the observed environment. From this perspective, it is clear that sensor and motor systems must closely function together to support perception and consequently, from a developmental point of view, must also evolve together. In this work we follow this line of thinking and propose an approach where sensor and motor structures develop conjointly into a well concerted sensorimotor system.

1.1 Related Work

The exploration of the advantages of temporally extended sensorimotor loops for perception and their implementation in artificial agents is still on-going work. On the one hand, sensorimotor learning traditionally focuses on learning a generally nonlinear coordinate transformation from a sensor related reference frame to motor space such that sensory input can be translated into a motor action appropriate for a task at hand [10, 14]. Moreover, with the advent of motor theories of perception [2, 3], a mapping of sensor and motor systems in the opposite direction – i.e. forward models predicting stimuli from motor commands – has gained attention in the robotics community too [18]. On the other hand, work which also considers structural adaptations of sensor and motor systems to obtain adequate sensorimotor maps, as addressed in this paper, is less widespread. This fact seems opposed to the importance of the induced coupling between a given motor apparatus and the physical structure of a sensory system which has been described by roboticists early on; e.g. for visual perception in [4], but also in general for embodied agents with different sensor modalities in [13]. A notable exception is a structurally adaptive sensorimotor system described in [8]. There a robot evolves a 1-dimensional visual sensor such that projected stimuli undergo a uniform translation during straight locomotion.

In a broader context, [9] analyses the causal structure present in the information flow induced by sensorimotor activity using information theoretic measures. The results in essence confirm that the characteristics of recorded stimuli have strong

ties to spatiotemporal relationships defined by the physical embodiment and the movement strategies executed by the considered artificial agent.

The authors of this article investigated in previous work the structure of linear stimulus prediction models for visual sensors [15]. It was found that the pairing of a particular sensor topology and sensor actuation strategy has a profound impact on the complexity of a visual stimulus prediction model. The adaptation of motor primitives with respect to a given visual sensor topology and, vice versa, the adaptation of a sensor topology given a particular interaction pattern were previously addressed in [16] and [17] considering these two problems independently.

1.2 Contribution

In this work, we develop a computational approach to conjointly synthesize visual sensor topologies and visual motor layers according to a given agent-environment interaction. The presented method takes as input an agent’s interaction pattern with its environment and evolves a spatial layout for both, light receptive fields and motor movement fields. The resulting sensorimotor structure is tuned to the characteristics of the agent’s interaction with its environment. We show that visual receptive fields and motor movement fields can evolve simultaneously when minimizing a simple error measure which contemplates the reconstruction error for recorded stimuli with respect to given input signals, and the prediction error for stimuli resulting from self-initiated actions. Driven by the predominantly low spatial frequency of natural images, spatially coherent and smoothly overlapping receptive fields organize on the sensor side without any further constraint on spatial shape. At the same time on the motor side, individual movement fields evolve such as to displace the sensor ensuring high temporal coherence of visual stimuli. Compared to [16], we additionally relax here the constraint that movement fields must implement a Gaussian model and instead allow them to evolve freely. At the beginning of the adaptation process, both, visual receptive fields and motor movement fields can be initialized with randomly chosen activation functions and eventually develop into compact fields.

2 Self-Organization of Visual Sensorimotor Structures

A common line of thinking in biology proposes that evolutionary adaptation implicitly optimizes some underlying criterion which is related to the fitness of an organism [12]. From an abstract perspective, it can be argued that similarly any autonomous artificial system should optimize a certain overall cost function in order to temporally maximize its resource-efficiency, task completion rate, or in general its functional subsistence. In the remainder of this work, we consider an artificial agent inhabiting a given world or ecological niche (N) developing so as to optimize an underlying cost function c_{agent} . Clearly, the function c_{agent} strongly depends on the

agent’s body and behavior, where the body of the agent can be further decomposed into its perceptual abilities (S) and its motor apparatus (M). The behavior (Q) is defined as a lifelong sequence of motor actions which depends on the agent’s particular survival strategy. We propose that a developmental process for the considered artificial agent should implicitly strive to optimize a loosely defined optimization problem

$$\min_{(S,M,Q)} c_{\text{agent}}(S, M, Q; N), \quad (1)$$

which can always be separated into

$$\min_Q \left[\min_{(S,M)} c_{\text{agent}}(S, M, Q; N) \right]. \quad (2)$$

Note that in this form, the full problem can be locally solved by iteratively optimizing first variables S and M while keeping Q constant and then optimizing Q while keeping S and M constant. Here, we are interested in optimizing sensorimotor structures (S, M) , and hence we address only the inner problem in (2) and consider Q fixed. In this case, the agent’s interaction with its environment can be recorded as a set of efferent and afferent signals experienced during lifetime according to Q .¹ In line with observations made in living organisms, the first hypothesis in this work is that the characteristics of such lifelong sensorimotor activity is the principal driving force for the co-development of sensorimotor structures (S, M) . With this hypothesis the inner optimization problem given in (2) can be rewritten as

$$\begin{aligned} \min_{(S,M)} c_{\text{sm}}(S, M; I_0, I_1, a) \quad , \\ \text{s.t. } (I_0, I_1, a) \sim B(Q, N) \end{aligned} \quad (3)$$

where the agent’s behavior Q and environment N enter the problem as overall experienced before-and-after signals (I_0, I_1) when executing actions a . The function B defines how triplets (I_0, I_1, a) are sampled from Q and N .

In problem (3), S and M describe the agent’s sensorimotor structure. Again, from an abstract perspective, both, sensory and motor systems can be considered a physical implementation which reduces in a specific way the dimensionality of perceivable stimuli and possible actions. In this sense, S can be thought of as a descriptor of the sensor’s structure which defines how the agent records a stimulus from available signals I . For visual sensors, such a structure is typically implemented as a 2-dimensional spatially non-uniform distribution of light sensitive receptors which linearly integrate luminance through receptive fields. In motor systems, a reduction in dimensionality can be considered to be present when lower level actions are organized into directly addressable higher level movements with some added value for the acting agent. For an example on how such a reduction in dimensionality is implemented in a biological system, see e.g. [6]. In this work, we address such a

¹ In neuroscientific terms, a motor signal sent from the central nervous system to the periphery of an organism is called *efference*. Conversely, a sensory signal traveling from the periphery of an organism to the central nervous system is called *afference*.

reduction in dimensionality for a very early motor layer. Similarly as for the sensory system, the structure of the motor system M is composed of discrete motor movement fields covering the available motor space non-uniformly in a way which provides an advantage for the considered agent. An example of such first layers of motor structures in biology are the motor layers in the optic tectum or superior colliculus as found in mammalian species [7].

With S and M encoding the structure of the agent’s sensorimotor system, we now incorporate the second hypothesis of this paper which addresses the co-development of S and M . We propose that sensory and motor systems organize so as to minimize the expected error between available signals I and stimuli which the agent actually records via $S(I)$. Reducing such an error directly relates to the request for the sensorimotor system to optimize available resources in favor of accurate perception. To measure a distance between $S(I)$ and I , a reconstructed signal $S^+(S(I))$ is compared to the original signal I , where S^+ projects the low-dimensional signal back into the original sensor space. Furthermore, for the perceptual process to work as a continuous sensorimotor loop, we not only want an accurate spatial relationship of the agent to its environment, but also maintain this relationship in a coherent manner over time. We thus include a coupling of sensory and motor systems via a prediction mechanism (P) capable of predicting future sensory stimuli from executed motor actions a .² Hence, the second hypothesis proposes that c_{sm} is of the form

$$\min_{(S,M,P)} \mathbb{E} \left[\left\| S^+(P(M,a,S(I_0))) - I_1 \right\|^2 \right], \quad (4)$$

s.t. $(I_0, I_1, a) \sim B(Q, N)$

where the norm is used to measure the reconstruction error of a predicted stimulus and the actually experienced signal. Other distance measures could be considered instead. The interested reader can find an excellent review on the ubiquity of stimulus prediction in living organisms e.g. in [3].

2.1 Realization

To solve problem (4), sensor and motor spaces are discretized as regular grids which yield sensor signals and motor activity as vectors \mathbf{I}_0 , \mathbf{I}_1 , and \mathbf{a} , sampled according to $B(Q, N)$. Note that similarly, as \mathbf{I} represents recorded activation on the given sensor surface, a motor action \mathbf{a} is a vector describing an activation profile on the motor space. Here it is assumed that the considered agent possesses a given motor system which transforms activation profiles \mathbf{a} into specific motor actions. Such a transformation can e.g. be thought to be a weighted vector sum of activated locations in the motor space, compare also [7] for an example in biological systems.

² Note that for a complex agent, the consequences of an action a might depend on the current state of the agent in which case the predictor P must be state aware. In this paper, as described in Sect. 2.1, we consider only cases where prediction is state independent.

Sensor and motor structures S and M are represented as positive matrices (\mathbf{S}, \mathbf{M}) which when applied to \mathbf{I} and \mathbf{a} yield visual stimuli $\mathbf{S}\mathbf{I}$ and motor movement field activations $\mathbf{M}^\top \mathbf{a}$. We choose \mathbf{S} and \mathbf{M} to be positive since on the sensor side \mathbf{S} represents light integrating receptive fields, and \mathbf{M} on the motor side encodes movement fields integrating activation from the underlying motor space. To predict sensory stimuli, we consider P to be a stateless predictor represented as a mixture of linear predictors of the form $P(\mathbf{M}, \mathbf{a}, \mathbf{S}\mathbf{I}_0) = [\sum_i \lambda_i(\mathbf{M}, \mathbf{a}) \mathbf{P}_i] \mathbf{S}\mathbf{I}_0$, as introduced in [16]. Additionally, we relax the constraint that λ_i must be composed solely of Gaussian receptive fields as in the previous work and instead allow for arbitrary field shapes $\lambda_i(\mathbf{M}, \mathbf{a}) = \mathbf{m}_i^\top \mathbf{a}$, where \mathbf{m}_i is the i -th column of \mathbf{M} . After prediction, the signal is reconstructed using the adjoint operator \mathbf{S}^\top . In this sense, we rewrite (4) as

$$(\mathbf{S}^*, \mathbf{M}^*, \mathbf{P}^*) = \underset{(\mathbf{S}, \mathbf{M}, \mathbf{P})}{\operatorname{argmin}} \sum_a \|\mathbf{S}^\top \sum_i [(\mathbf{m}_i^\top \mathbf{a}) \mathbf{P}_i] \mathbf{S}\mathbf{I}_0 - \mathbf{I}_1\|^2. \quad (5)$$

s.t. $\mathbf{S} \geq \mathbf{0}, \quad \mathbf{M} \geq \mathbf{0}, \quad \mathbf{P} \geq \mathbf{0}$

The savvy reader will notice that the apparent ambiguity which arises by the interaction between \mathbf{P} and \mathbf{M} nearly disappears with the positivity constraints.

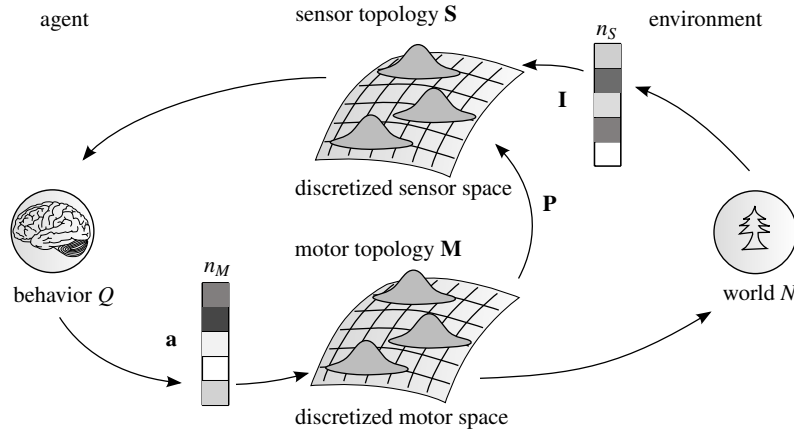


Fig. 1 The sensorimotor loop considered when organizing lower-dimensional sensory and motor topologies \mathbf{S}, \mathbf{M} and learning the stimulus predictor \mathbf{P} . On the motor side, the agent generates motor commands \mathbf{a} of size n_M according to a given behavior Q . On the sensor side, the agent experiences input signals \mathbf{I} of size n_S which represent a projection of the world onto the sensor. When executing the full sensorimotor loop, each action \mathbf{a} changes the input signal \mathbf{I} and generates a triplet $(\mathbf{I}_0, \mathbf{I}_1, \mathbf{a})$. During learning, the lower-dimensional sensor and motor topologies \mathbf{S} and \mathbf{M} evolve according to given triplets $(\mathbf{I}_0, \mathbf{I}_1, \mathbf{a})$. At the same time, the prediction operator \mathbf{P} is learned such as to predict future sensory stimuli $\mathbf{S}\mathbf{I}_1$ from previous stimuli $\mathbf{S}\mathbf{I}_0$ for any action \mathbf{a} .

2.2 Method

We consider the organization of $N_S = 16$ visual receptive fields taking place on a sensor surface in the shape of a disk discretized at $n_S = 481$ locations in a grid-like layout. Similarly, experiments presented in Sect. 3 consider $N_M = 16$ motor movement fields evolving on 2-dimensional motor spaces discretized at $n_M = 15 \times 15$ locations in a grid-like layout, see also Fig. 2. The environment is given as a plane textured by a very high resolution image (2448×2448 pixels) depicting a real world scene. In this article we assume the sensor surface to be parallel to the plane recording grayscale images \mathbf{I} . The sensor can interact with the environment through four types of actions, translations in x- and y-directions, rotations and changes in distance to the plane (zoom). A set of 22500 triplets $(\mathbf{I}_0, \mathbf{I}_1, \mathbf{a})$ is obtained via $B(Q, N)$, where for the presented experiments the underlying Q selects actions \mathbf{a} with sharp activation profiles (all entries in \mathbf{a} are zero except one) according to a uniform distribution over the discretized action space. Each triplet is obtained by positioning the agent in a random position on the environment and taking the chosen action a . To find $(\mathbf{S}^*, \mathbf{M}^*, \mathbf{P}^*)$, we iteratively improve the optimization problem given in Eq. (5) using a projected gradient descent method [1]. While it is no problem to find a solution with an online method, convergence is much slower, we therefore choose here the batch approach for practical reasons. However, we note that under different circumstances an online implementation might be preferable, e.g. for a purely biologically inspired implementation in a robot with stronger memory constraints and a longer exploration phase. The experiments presented in Sect. 3 were initialized as follows: the motor layout \mathbf{M} randomly according to a uniform distribution between zero and one; \mathbf{S} randomly such that each discrete sensor location belongs to exactly one receptive field (row of \mathbf{S}), scaled so as to obey $\mathbf{S}\mathbf{S}^\top = \mathbb{I}$. The prediction matrices \mathbf{P}_i were initialized with given random \mathbf{S} and \mathbf{M} to the least squares solution to predict $\mathbf{S}\mathbf{I}_1$ with $[\sum_i \lambda_i (\mathbf{M}, \mathbf{a}) \mathbf{P}_i] \mathbf{S}\mathbf{I}_0$ and subsequently projected according to $\mathbf{P} \geq \mathbf{0}$. It is important to note that with a randomized initialization, nothing prevents the adaptation process from converging to a locally optimal solution. However, from a biological point of view, we accept these solutions as possible branches of evolutionary development.

2.3 Implication

The presented approach for the co-development of visual sensor and motor structures is based on two main hypotheses. The first states that sensorimotor structures can be decoupled within problem (2) in the sense of (3), and the second proposes that S and M evolve such as to optimize i) the reconstruction of higher dimensional signals, and ii) stimulus predictability. Per se, it is not clear if these hypotheses are justifiable. However, if the proposed framework is capable of reproducing some characteristics of in nature observed sensorimotor structures, then an indication is provided that the implementation captures some inherent principles present in phy-

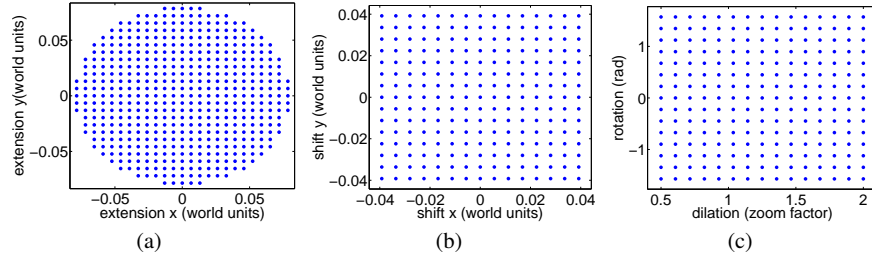


Fig. 2 (a) Discretization of the given sensor space; (b) discretization of the given motor space for a behavior with horizontal and vertical translation actions; (c) discretization of the given motor space for a behavior with dilation and rotation actions. Sensor area and translation distances are specified in world coordinates ranging from -1 to 1 in x- and y-direction.

logenetic and or ontogenetic development of biological systems. In this case, even though we are not aware of the true evolutionary cost function, we might claim that the made assumptions could hold, and that the proposed framework with its simple underlying principles has explanatory power.

3 Results

Two different behaviors Q_1 and Q_2 were considered to co-develop sensor and motor topologies \mathbf{S}_1^* , \mathbf{S}_2^* and \mathbf{M}_1^* , \mathbf{M}_2^* . In a first setup, $B(Q_1, N)$ samples sensor translation actions from a 2-dimensional motor space of a given range as shown in Fig. 2(b). Triplets $(\mathbf{I}_0, \mathbf{I}_1, \mathbf{a})$ are sampled choosing actions \mathbf{a} with uniform probability from the available discrete actions. This scenario relates to translational unbiased oculomotor control causing random stimulus displacements. The second behavior is composed of mixed zoom and rotation actions where $B(Q_2, N)$ samples combined sensor rotations and stimulus dilations from a 2-dimensional motor space as shown in Fig. 2(c). As for Q_1 , triplets $(\mathbf{I}_0, \mathbf{I}_1, \mathbf{a})$ were sampled with uniform probability from the available discrete actions. Behavior Q_2 mimics an object manipulating agent where the oculomotor system stabilizes the sensor on target, mechanically compensating for image translations but not image rotations or scaling. The resulting sensor and motor topologies \mathbf{S}_1 , \mathbf{S}_2 and \mathbf{M}_1 , \mathbf{M}_2 are shown in Fig. 3. The results demonstrate that different behaviors Q induce sensorimotor structures of different macroscopic nature. Note that even though the proposed algorithm is unaware of the topological order present in recorded stimuli \mathbf{I} , visual receptors cluster as smoothly overlapping receptive fields and motor primitives appear as spatially coherent Gaussian-like areas.

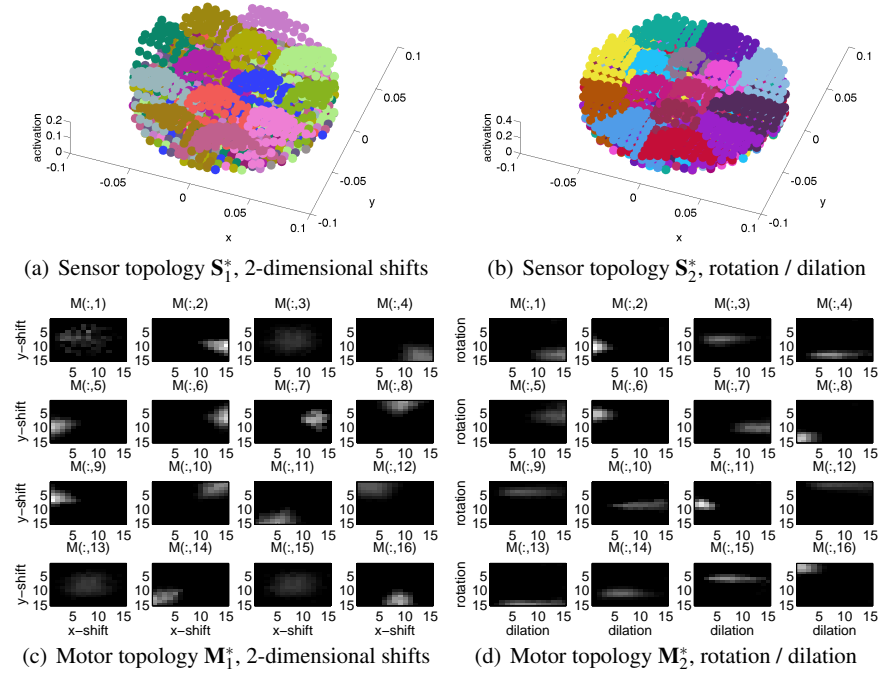


Fig. 3 Sensor and motor topologies obtained under behaviors with actions uniformly sampled from motor spaces as shown in Fig. 2(b) and Fig. 2(c). Resulting sensor layouts are shown in (a) and (b) where each color denotes a different visual receptive field, and each dot shows the activation of that field at the respective location on the sensor surface. In the translation only case we can identify a tendency for hexagonal tiling structures, whereas in the rotation and dilation case the receptors organize more radially. In (c) and (d) the evolved motor movement fields are shown. For the translation only case, motor fields organize as compact Gaussian-like areas, whereas in the rotation and dilation case, elongated elliptical fields develop reflecting the higher axial resolution of sensor S_2^* compared to its radial resolution. Note that some motor fields happen to overlap and therefore appear less pronounced as their contribution is combined according to Eq. (5).

4 Conclusions and Outlook

This paper investigated how the behavior of an artificial agent can shape the sensorimotor structure of its visual system. We proposed that well adapted sensor and motor layouts organize such as to accurately represent given input signals not only spatially but also temporally for a set of motor actions characteristic for the behavior of the considered agent. We showed that this criterion is captured by comparing the reconstruction of a predicted future stimulus and the actually experienced signal and can be used to conjointly develop visual receptive fields and motor movement

fields. In living organisms, comparable structures mapping visual sensory input to motor output can be found in the optic tectum or superior colliculus in mammals.

In future work, we intend to address larger scale problems with an optimized version of the current implementation, which eventually could also serve as a design tool for synthesizing behavior-specific sensorimotor structures for artificial agents. Furthermore, we plan to apply the introduced principles to other sensory modalities, e.g. in a frequency domain for auditory perception.

Acknowledgements This work was supported by the European Commission proj. FP7-ICT-248366 RoboSoM, by the Portuguese Government – Fundação para a Ciência e Tecnologia (FCT) proj. PEst-OE/EEI/LA0009/2011, proj. DCCAL PTDC/EEA-CRO/105413/2008, and FCT grant SFRH/BD/44649/2008.

References

1. Absil, P.A., Mahoney, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press (2008)
2. Craighero, L., Fadiga, L., Rizzolatti, G., Umiltà, C.: Action for perception: A motor-visual attentional effect. *Exp. Psych.: Human Perception and Performance* **25**, 1673–1692 (1999)
3. Crapse, T.B., Sommer, M.A.: Corollary discharge across the animal kingdom. *Nat. Rev. Neuroscience* **9**, 587–600 (2008)
4. Franceschini, N., Pichon, J.M., Blanes, C.: From insect vision to robot vision. *Phil. Trans. Biological Sciences* **337**, 283–294 (1992)
5. Gibson, J.J.: The theory of affordances. In Shaw et al., *Perceiving, Acting, and Knowing: Toward an Ecological Psychology* pp. 67–82 (1977)
6. Kazuya, S., Ménard, A., Grillner, S.: Tectal control of locomotion, steering, and eye movements in lamprey. *J Neurophysiol.* **97**, 3093–3108 (2007)
7. Lee, C., Rohrer, W.H., Sparks, D.L.: Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature* **332**, 357–360 (1988)
8. Lichtensteiger, L., Eggenberger, P.: Evolving the morphology of a compound eye on a robot. In: *Proc. 3rd Europ. Worksh. on Adv. Mobile Robots*, pp. 127–134 (1999)
9. Lungarella, M., Sporns, O.: Mapping information flow in sensorimotor networks. *PLoS Computational Biology* **2**, 1301–1312 (2006)
10. Massone, L.L.E.: Sensorimotor learning. In: M.A. Arbib (ed.) *The Handbook of Brain Theory and Neural Networks*, pp. 860–864. The MIT Press (1995)
11. O’Regan, J.K., Noë, A.: A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* **24**(5), 939–1031 (2001)
12. Parker, G.A., Smith, J.M.: Optimality theory in evolutionary biology. *Nature* **348**, 27–33 (1990)
13. Pfeifer, R., Bongard, J.: *How the Body Shapes the Way We Think*. MIT Press (2006)
14. Pouget, A., Snyder, L.: Computational approaches to sensorimotor transformations. *Nature Neuroscience* **3**, 1192–1198 (2000)
15. Ruesch, J., Ferreira, R., Bernardino, A.: A measure of good motor actions for active visual perception. In: *Proc. Int. Conf. on Development and Learning (ICDL)*, Frankfurt, Germany. IEEE (2011)
16. Ruesch, J., Ferreira, R., Bernardino, A.: Predicting visual stimuli from self-induced actions: an adaptive model of a corollary discharge circuit. *IEEE Transactions on Autonomous Mental Development*, accepted (2012)
17. Ruesch, J., Ferreira, R., Bernardino, A.: Self-organization of visual sensor topologies based on spatio-temporal cross-correlation. In: *Proc. 12th Int. Conf. on Adaptive Behavior (SAB)*, Odense, Denmark, accepted. Springer (2012)
18. Wolpert, D.M., Diedrichsen, J., Flanagan, J.R.: Principles of sensorimotor learning. *Nature Reviews Neuroscience* **12**, 739–751 (2011)