Adaptive Behavior
000(00):0–0
©The Author(s) 2010
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI:doi number
http://mms.sagepub.com

A computational approach on the co-development of artificial visual sensorimotor structures

Jonas Ruesch and Ricardo Ferreira and Alexandre Bernardino*
Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon, Portugal

Abstract

To follow a goal-directed behavior, an autonomous agent must be able to acquire knowledge about the causality between its motor actions and corresponding sensory feedback. Since the complexity of such sensorimotor relationships directly influences required cognitive resources, this work proposes that it is of importance to keep the agent's sensorimotor relationships simple. This implies that the agent should be designed in a way such that sensory consequences can be described and predicted in a simplified manner. Living organisms implement this paradigm by adapting sensory and motor systems specifically to their behavior and environment. As a result, they are able to predict sensorimotor consequences with a strongly limited amount of (expensive) nervous tissue. In this context, the present work proposes that advantageous artificial sensory and motor layouts can be evolved by rewarding the ability to predict self-induced stimuli through simple sensorimotor relationships. Experiments consider a simulated agent recording realistic visual stimuli from natural images. The obtained results demonstrate the ability of the proposed method to i) synthesize visual sensorimotor structures adapted to an agent's environment and behavior, and ii) serve as a computational model for testing hypotheses regarding the development of biological visual sensorimotor systems.

Keywords

sensorimotor learning, sensorimotor coupling, self-organization of sensorimotor structures, visual receptors, visual motor primitives, visual stimulus prediction.

1. Introduction

In a situation where an artificial system fails to select a sensible action despite the potential availability of required sensory information, one could claim the system lacks the required cognitive skills to find a solution. However, conversely it can be argued, the reason why the systems fails to deduce an appropriate action is simply because the relationship between sensory and motor signals is too complicated in order for the robot to translate recorded sensory stimuli into a motor action adequate in the current context. While the first formulation suggests to increase the robot's "brain power", the second formulation implies an alternative approach. Instead of striving to increase cognitive capabilities, one could try to find a design for the robot's sensorimotor apparatus such that less complex operations are required to translate sensory feedback into purposeful actions. Seen from the latter perspective it can be conjectured: if sensory and motor systems of a robot are well concerted and adapted to a specific behavior and environment, then the cognitive load imposed on the agent's processing system is reduced.

This view that appropriate sensorimotor morphology can save cognitive resources, is also strongly supported by evolutionary biology and research in neuroscience. A number of studies showed that for biological systems there is high selective pressure to evolve specialized sensorimotor systems which allow for simpler nervous systems and enable efficient solutions for stimulus processing. For an overview see for example work by Niven & Laughlin (2008), Olshausen & Field (2004), Chklovskii (2004), Vinje & Gallant (2000), and Cherniak (1995). Thus, considering the costs associated with the development of neural tissue and studies demonstrating the degeneration of neural structures under relieved selective pressure, it is reasonable to assume that as long as a problem can be solved by "cheaper" means – like evolving a more suitable sensorimotor system – the development of more resource-intensive cognitive abilities is delayed in biological systems.[1] According to this line of thinking, this work considers the following two points of crucial importance for the design of autonomous agents:

- The sensory system should provide stimuli which are meaningful with respect to the agent's motor capabilities and environment. Or, conversely, that sensing something which is never relevant with respect to a possible motor action is redundant.
- The motor system should favor actions that facilitate the prediction of their effects in the perceptual stream. Or conversely, actions which lead to complex or inaccurate predictions of percepts are unfavorable since they hamper the linkage between consecutive sensory stimuli and the ability to plan actions ahead in time.

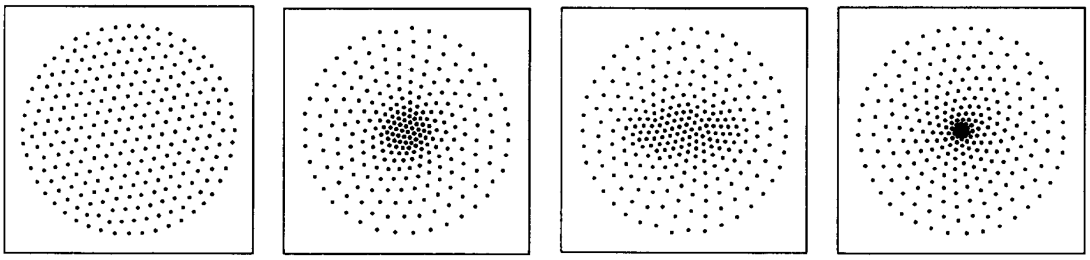
In robotics and artificial intelligence, the strategy of solving a problem with less computational power through a specifically designed body has sometimes been referred to as "morphological computation", see for example Paul (2006); Pfeifer et al. (2006). A great number of examples demonstrating this approach for arbitrary artificial systems have been described by Pfeifer & Scheier (1999) and Pfeifer & Bongard (2006). These examples show how the cognitive load imposed on artificial agents can be reduced by taking advantage of the morphological characteristics of an agent's body and the properties of the ecological niche inhabited. Lichtensteiger & Eggenberger (1999) for example describe a robot with an adjustable 1-dimensional visual sensor which learns to change the distribution of its visual receptors such that projected stimuli undergo a uniform translation during straight locomotion. The proposed optimization of the sensor relates to the idea of reducing cognitive resources through structural changes in the sense that the resulting receptor distribution facilitates visual distance estimation. Other examples of artificial systems which make use of "cheap" visual perception include robots with solve complex tasks like navigation, visual tracking, or object recognition (Wehner, 1987; Franceschini et al., 1992, 2007; Reiser & Dickinson, 2003; Floreano et al., 2004).

1.1. Self-similar sensorimotor structures

A particularly inspiring work with respect to the organization of sensor topologies has been published by Clippingdale & Wilson (1996). Due to the importance of the concepts brought forward by this work to the comprehension of our approach, this section provides a dedicated review.

Clippingdale and Wilson propose to organize abstract representations of sensor topologies by maximizing a measure for the sensor's self-similarity under a given set of transformations. A sensor is considered self-similar if there exist motor actions which move the sensor in a way such that recorded visual stimuli are dispaced exactly or approximately from receptor to receptor. In this context, the work by Clippingdale shows that a set of points, initially randomly distributed on a planar disk, converges to a stable configuration with a highly regular structure under the following rules: i) points are conjointly transformed by rotation, dilation, and translation actions which are applied according to a given probability distribution, and ii) after each iteration points are moved towards transformed points lying closest. Interestingly, it was found that under the described conditions, certain action probability distributions induce point distributions which resemble closely receptor topologies found in foveal sensor layouts of camera-type visual systems. The results of Clippingdale and Wilson are shown in Fig. 1. The probability distributions of transformations which lead to the configurations shown in these figures are composed of

[1] Of course, to what extent cognitive resources can be "saved" by implementing smart adaptations depends on the specific problem to be solved and the associated potential for specialization. Furthermore, a highly adapted design always comes at the cost of loosing the ability to address more general problems.



(a) Trans. h = 1.0, v = 1.0 (b) Trans. h = 0.19, v = 0.2 (c) Trans. h = 0.5, v = 0.2 (d) Trans. h = 0.0, v = 0.0

Fig. 1. Geometrical point layouts obtained by the algorithm described by Clippingdale & Wilson (1996). The different plots show layouts obtained under different layout transformation action probability distributions. Actions are sampled from a uniform distribution of arbitrary rotation and dilation actions, and a uniform distribution of x- and y-translations of limited range. The different ranges of translation actions are denoted below each plot. Reproduction with the kind permission of S. Clippingdale.

rotation and dilation transformations uniformly distributed over an arbitrary range, combined with horizontal and vertical translations distributed over different limited ranges.[2] In summary, Clippingdale and Wilson show that abstract sensor layouts can be synthesized which resemble receptor distributions found in biological organisms and, that these layouts are obtained under stimulus transformations which can be expected to be characteristic for the respective organisms. It is important to note that the measure for self-similarity, as introduced by Clippingdale and Wilson, directly relates to a request for simpler feedforward sensorimotor maps. It essentially favors (on average) exact receptor-to-receptor stimulus displacements.

A number of restrictions related to the approach proposed by Clippingdale & Wilson (1996) are noteworthy in the context of this work. Firstly, it is assumed that there is knowledge about the sensor topology, meaning the algorithm has access to the spatial position of sensory elements. Secondly, it is assumed that the displacement of sensor elements with respect to the sensor surface is known a priori from a given transformation; which for a real visual sensor has to be reformulated as: it is assumed that the new locations of a previously recorded stimuli are known from a given transformation. Both assumptions are unrealistic when considering an autonomously developing organism or robot. Rather, it seems appropriate to take an intrinsic perspective of a developing system and to base self-organization of the sensorimotor apparatus on recorded stimuli and motor commands directly available to the system. From this point of view, it cannot be assumed that measurements concerning the spatial topology of the sensory system can be obtained. Also, it is improbable that information about the spatial displacement of visual receptors is readily available since motor commands and stimulus displacements are usually connected via complex transformations. In contrast, our work adopts a more natural approach where the sensor topology is considered unknown and visual stimuli are recorded with a realistic sensor model.

1.2. Contribution

The principal goal of the present work is to develop a method which allows for the self-organized synthesis of sensorimotor structures in artificial visual systems. Considering a co-developmental process, it is proposed that sensory and motor systems mutually influence each other such that the structure of a visual sensor organizes according to the characteristics of the given motor apparatus, and, vice versa, the motor system adapts to the structure of the associated sensor. As a common direction for such a joint development, the present work proposes that sensor morphology and motor primitives should concurrently adapt such as to simplify the relationship between recorded sensory stimuli and executed motor actions. A more concrete indication on how to qualify "simple" sensorimotor relationships in visual systems has been previously introduced in Sect. 1.1 through the concept of self-similarity from Clippingdale & Wilson (1996).

In Fig. 2, a sketch of a general sensorimotor system is shown. Dashed lines indicate components of the system

[2] In Clippingdale & Wilson (1996) Clippingdale and Wilson extended the same algorithm to work with 2-dimensional Gaussians instead of simple geometrical points. To do so, the Euclidean distance measure between two points was replaced with the inner product between two Gaussians in the self-similarity measure.

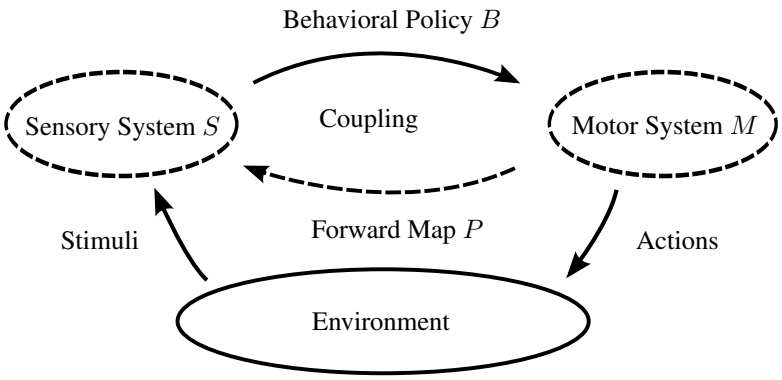


Fig. 2. A general sensorimotor system with subsystems S and M which are coupled in two directions via sensorimotor maps B and P. Dashed lines denote initially unknown components of the system. The policy B associates actions to recorded stimuli and is considered to implement a given behavior. The sensorimotor system (S, M) and its sensorimotor transformation (P) are variables of the problem.

which are initially unknown. Thus, both, the sensory system S, and the motor system M, are considered variables of the proposed problem. The two arrows between S and M indicate the two directions of possible sensorimotor transformations. The arrow denoted behavioral policy B represents the map which induces a temporal distribution of actions. This action distribution is always considered to be known throughout this work, as in (Clippingdale & Wilson, 1996). In the opposite direction, the arrow denoted forward model P represents the sensorimotor map which estimates sensory stimuli based on selected motor actions. This map is initially unknown and represents a variable of the problem which depends on S and M. Having proposed that a perceptual system with favorable properties possesses a simplified feedforward model, the structure of P is of central importance. Accordingly, the central idea is to adapt S and M such that a simplified feedforward model P results. The adaptation of the structure of an internal feedforward model is plausible for a biologically inspired system since forward models for stimulus prediction are ubiquitous in living organisms. Neural circuits concerned with the prediction of sensory stimuli from motor signals are in general termed corollary discharge circuits. An excellent overview of corollary discharge circuits is provided e.g. by Crapse & Sommer (2008). Other work describing and modeling biologically inspired feedforward sensorimotor maps has been described by Wolpert & Kawato (1998); Webb (2004); Wolpert et al. (2011).

The presented method takes as input experienced efferent (motor commands) and afferent (stimuli) signals, and evolves a spatial layout for light receptive fields and motor movement fields as well as the prediction model. It is shown that visual receptive fields and motor movement fields can evolve simultaneously when minimizing a simple error measure which contemplates the prediction error for stimuli resulting from self-initiated actions. On the sensor side, the low spatial frequency of natural images induces the development of spatially coherent and smoothly overlapping receptive fields without any further constraint on spatial shape. At the same time on the motor side, individual movement fields evolve such as to encode actions ensuring high temporal coherence of visual stimuli. The presented results demonstrate how the proposed principles can be used to develop sensory and motor systems with favorable mutual interdependencies.

A constraint which is imposed throughout the present work is that sensorimotor relationships are independent of the agent's state; i.e. changes in visual stimuli only depend on the taken motor actions and not for example on the system's current position with respect to the environment. This is a good approximation, for instance, for a visual agent that moves in an environment where objects are far away with respect to the motion amplitudes (low parallax effects).

2. Problem formulation

This work considers an artificial agent consisting of a sensor S, composed of an array of light sensitive receptors s, and a motor layer M, composed of a number of motor primitives m. The agent lives in a static environment and works as

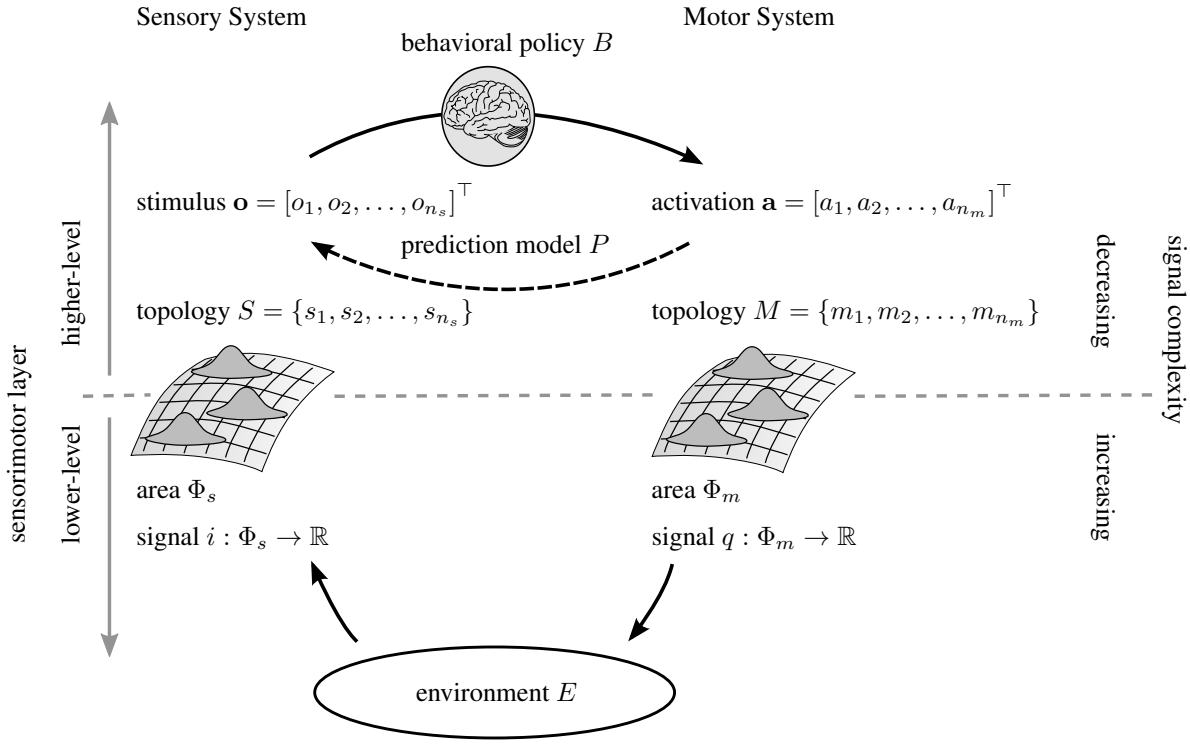


Fig. 3. Schematic overview of the considered sensorimotor system with the sensory pathway on the left and the motor pathway on the right. The figure illustrates lower-level sensorimotor layers as grids and higher-level sensorimotor layers as Gaussian shaped receptive fields.

a closed sensorimotor loop. It observes the stimulation of its receptor array and takes actions by activating its motor primitives depending on the recorded stimulus. To choose actions based on recorded stimuli the agent possesses a given behavioral policy B which associates to any given stimulus an action. Thus, for a sequence of discrete time steps, the agent records at each time step a stimulus as the activation of its receptors $\mathbf{o}$ and triggers an action by activating its array of motor primitives via weights $\mathbf{a}$. An action $\mathbf{a}$ can induce a change in the observed stimulus which leads to the selection of a new action in the next time step according to the policy B. Stimuli $\mathbf{o}$ resulting after an action $\mathbf{a}$ is selected are estimated by the agent through the sensorimotor map P (initially unknown). A complete schematic overview of the discussed sensorimotor system is given in Fig. 3.

The agent's body physically implements sensory and motor systems in areas Φ_s and Φ_m. These areas are assumed to be given and represent the physical space where sensory and motor signals are recorded and encoded. In a biological system, Φ_s corresponds for example to the surface of a retina recording a projection of the environment. Similarly on the motor side, Φ_m corresponds to a layer or volume of neural tissue where each location represents a particular motor signal. Hence, considering a d_s-dimensional sensor, the sensor area Φ_s is a topographic map of $\mathbb{R}^{d_s}$ and represents the domain on which visual signals $i : \Phi_s \to \mathbb{R}$ are defined. Analogously, for an agent with d_m degrees of freedom, the motor area Φ_m is a topographic map of $\mathbb{R}^{d_m}$ and represents the domain on which motor signals $q : \Phi_m \to \mathbb{R}$ are defined.[3] In the remainder of this work, however, the morphology of Φ_s and Φ_m can be abstracted. Of principal importance is the existence of visual signals $i \in \mathcal{I}$ and motor signals $q \in \mathcal{Q}$ contained in the agent's sensor space $\mathcal{I}$ and motor space $\mathcal{Q}$. Still, the fact that Φ_s and Φ_m are not explicitly considered, does not mean that their structure has no influence on the organized sensorimotor system, rather, their physical shape influences how signals i and q are recorded, and these signals in turn influence the system processing i and q. Also, the assumption that Φ_s and Φ_m are

[3] In a physical implementation the motor area Φ_m is at most a volume. Thus, if $d_m > 3$, a mapping of $\mathbb{R}^{d_m}$ to Φ_m is required. Such mappings of higher dimensional spaces to volumes or layers of neural tissue is commonly found in biological systems. For a model see e.g. Swindale (2004).

given, implies that part of the agent's morphology is predefined. On the sensor side, this concerns physical structures used to form an image of the environment on the sensor. For a camera-type eye, this is for example a lens system and the shape of the retinal surface. On the motor side, actions are encoded in Φ_m which essentially means it is assumed that the kinematics of the considered agent is given.

The agent records visual signals i through a layer of n_s light receptive elements $S = [s_1, s_2, \ldots, s_{n_s}]$ where each $s_j > 0$ represents a positive function encoding the integration area of the respective receptor. Likewise, the agent encodes motor signals q using a layer of n_m motor primitves $M = [m_1, m_2, \ldots, m_{n_m}]$ where each $m_k > 0$ represents a positive function encoding the influence area of the respective motor primitive. These layers define how visual signals i are reduced to compact visual stimuli $\mathbf{o}$ and how the activation of a limited number of motor primitives $\mathbf{a}$ composes and action q. The topological arrangement and physical shape of the n_s visual receptors and n_m motor primitives is initially completely unknown. Their structure represents the principal variable of the considered problem. A more in depth discussion on visual receptors and motor primitives and their biological prototypes follows below in this section. A forward sensorimotor map P, later used to drive the organization of S and M, is implicitly induced for any S and M. In general, a predictor P can be considered a functor which predicts the stimulus $\mathbf{o}_{t+1}$ for a given action q as $p\left(q, \mathbf{o}_t\right)$. A concrete prediction model suitable to implement p is presented later in Sect. 3.

2.1. Input to the problem

The input to the formulated problem is solely related to sensorimotor activity experienced by the considered agent. This activity is generated by the agent's behavioral policy which at each time step t selects an action q_t depending on the currently experienced stimulus. The behavioral policy B is assumed to be part of the given system and can be thought of as a simple brain of the agent, compare also Fig. 3. Thus, for an agent following a policy B, each time step a sensorimotor experience $e_t = (i_t, i_{t+1}, q_t)$ results, where i_t and i_{t+1} denote afferent signals recorded before and after the efferent signal q_t is elicited. The set of sensorimotor experiences $\mathcal{B} = \{e_t, t = 0 \ldots T\}$ is henceforth referred to as the sensorimotor experience of the agent.

An important hypothesis of this work is that the characteristic of an agent's sensorimotor interaction described by $\mathcal{B}$ is the principal driving force for the mutual development of sensorimotor structures S and M. Thus, $\mathcal{B}$ represents the input to the problem considered in this work.

2.2. Optimization

From an abstract perspective, it can be argued that autonomous and adaptive systems – no matter if they are of artificial or biological nature – should optimize a certain overall cost function in order to temporally maximize their resource-efficiency, task completion rate, or in general their functional subsistence (Parker & Smith, 1990). Thus, in this work it is assumed that the agent considered develops so as to optimize an underlying cost function c_{agent}. Clearly, such a function strongly depends on the agent's body and behavior, and with it on the structure of its sensorimotor apparatus (S, M). Here, it is proposed that a developmental process for the considered artificial agent should implicitly strive to optimize a loosely defined optimization problem

$$\min_{(S,M,B)} \; c_{\text{agent}}(S, M, B, E), \tag{1}$$

With the hypothesis that the agent's behavior B and environment E enter the problem as sensorimotor experience $\mathcal{B}$ as defined in the previous section, the inner optimization problem given in (1) can be rewritten as

$$(S^*, M^*) = \text{argmin}_{(S,M)}[c_{sm}\left(S, M; \mathcal{B}\right)]. \tag{2}$$

In this equation the predictor operator P is implicitly present because the sensorimotor structure (S, M) automatically induces a forward sensorimotor map. Thus, concerning the further elaboration of Eq. (2), it is clear that in order to incorporate the ideas outlined in Sect. 1, the cost function c_{sm} must be related to the accuracy and simplicity of the induced prediction model P.

3. Realization

3.1. Visual stimuli and motor corollary discharge

In order to represent the above described problem on a computer, continuous signals and functions have to be discretized. A general discretization of functions defined on continuous domains Φ_s and Φ_m is obtained by discretizing sensor and motor areas Φ_s and Φ_m in a grid-like manner. Thus, defining the resolution of discretized sensor and motor areas as N_s and N_m, visual and motor signals i and q can be represented as a real-valued vectors $\mathbf{i}$ and $\mathbf{q}$ of size N_s and N_m respectively. To represent motor signals q which are of the form of a Dirac delta function, a vector with a single non-zero entry denoting the location of the peak of the function is used. To represent receptive fields and movement fields in a discretized form, functions s and m are discretized accordingly as real-valued vectors $\mathbf{s}$ and $\mathbf{m}$. Sensor and motor topologies S and M can thus be represented as matrices $\mathbf{S}$ and $\mathbf{M}$ of size $N_s \times n_s$, respectively $n_m \times N_m$.[4] With this notation, the observation of stimuli $\mathbf{o}$ can be written as

$$\mathbf{o}_t = \mathbf{S} \cdot \mathbf{i}_t, \tag{3}$$

where $\mathbf{S}$ describes with each row a receptive field. The motor corollary discharge signal of a given action $\mathbf{q}$ is integrated by motor movement fields as

$$\tilde{\mathbf{a}} = \mathbf{M}^\top \cdot \mathbf{q}, \tag{4}$$

where $\mathbf{M}$ describes with each column a movement field and an estimation $\tilde{\mathbf{a}}$ of the action $\mathbf{a}$ which generated $\mathbf{q}$ is returned.

3.2. Stimulus prediction

A general stimulus predictor has been previously introduced as $p(q, \mathbf{o}_t)$. Also, it was anticipated that the complexity of p is related to the sensorimotor structure described by S and M. Consequently, if it is desired to organize S and M, such as to induce a simpler sensorimotor map, it is in the operator p where this request has to be accommodated. This section first proposes a concrete form for p, and subsequently outlines how to infer or enforce simplicity in p.

[4] The format of $\mathbf{M}$ was chosen to be transposed with respect to the format of $\mathbf{S}$. In this way, a natural application of $\mathbf{S}$ and $\mathbf{M}$ to $\mathbf{i}$ and $\mathbf{q}$ results.

Regarding the formulation of the prediction operator, the following is observed. Considering the static environment E and a spatially rigid sensor layout S, the class of functions from which a stimulus predictor p should be chosen can be restricted. In (Ruesch et al., 2012), an argument is provided which motivates a reduction of these functions to the linear function set.

$$\mathbf{o}_{t+1} \approx \mathbf{P}(q)\, \mathbf{o}_t, \tag{5}$$

where $\mathbf{P}(q)$ is the matrix representation of a linear prediction function $p(q)$.

To incorporate the tendency towards simpler prediction operators, it is now left to decide on the complexity of the prediction model. A common and natural approach to select simpler models is to evaluate the number of parameters required by the model. For a linear predictor, this requirement can be translated by forcing the predictor to be sparse. In this sense, equation Eq. (5) is revised as:

$$\mathbf{o}_{t+1} \approx \mathbf{P}(q)\, \mathbf{o}_t, \quad \mathbf{P}(q) \text{ sparse}. \tag{6}$$

This equation is still ill defined since the notion of sparsity is vague and nothing has been said about the prediction error. It is the authors' belief that these cannot be canonically defined, so several alternatives can be proposed as a means of mixing or balancing the importance of sparsity and allowed error:

- Fix sparsity and minimize some norm of the prediction error. For example one can say that for each location q in the function $\mathbf{P}(q)$, a matrix is found where each row of the matrix $\mathbf{P}^q$ must have a k non-zero entry (sparsity) and under this set the norm error is minimized.
- Minimize the prediction error and infer sparsity. A strategy which first obtains the minimum norm error solution for every location q in $\mathbf{P}(q)$ and subsequently deduces sparsity.
- Simultaneously minimize both prediction error and sparsity. For example the well known LASSO algorithm allows for a single parameter to weight the importance of sparsity versus norm error Tibshirani (1996).

Any of these methods will obtain prediction matrices $\mathbf{P}^q$ as well as associated prediction errors E^q from several samples of the sensor values before and after executing a specific action q from randomly chosen states x.

Following the second approach listed above, in (Ruesch et al., 2011) an empirical study is presented which investigates the sparsity of prediction operators resulting from the coupling of different sensor topologies and motor actions q. Interestingly, the obtained results indicate that for positive linear predictors $\mathbf{P}^q \geq \mathbf{0}$, the local maxima of the sparsity measure coincide with the local minima of the mean squared error E. These empirical results are a hint that the request for sparse predictors $\mathbf{P}^q \geq \mathbf{0}$ can be simply addressed through a minimization of E^q. The solution presented in the next section is based on this insight.

3.3. Proposed solution

To formulate a concrete instance of the problem outlined in Eq. (2) two aspects are incorporated. Firstly, sensory and motor systems should be coupled via the previously introduced stimulus prediction mechanism. And secondly, sensory and motor systems should organize so as to minimize the expected error between available signals $\mathbf{i}$ and stimuli which the agent actually records as $\mathbf{Si}$. Reducing this error directly relates to the request for the sensorimotor system to optimize available resources in favor of accurate perception.

In order to implement the temporal coupling of sensory and motor systems, the agent needs to implement $\mathbf{P}(\mathbf{q})$. Given an associated predictor $\mathbf{P}_k$ for each movement field $\mathbf{m}_k$, an approximation of the function $\mathbf{P}(\mathbf{q})$ can be constructed for actions $\mathbf{q}$ as the mixture of linear predictors $\mathbf{P}_k$ like

$$\mathbf{P}(\mathbf{M}, \mathbf{q}) = \sum_{k}^{n_m} \left({\mathbf{m}_k}^\top \mathbf{q} \right) \mathbf{P}_k, \tag{7}$$

where ${\mathbf{m}_k}^\top \mathbf{q}$ denotes the activation of a particular movement field. This model for visual stimulus prediction has been previously explored in (Ruesch et al., 2012) where also the accuracy of such a linear combination has been assessed.

To enable the comparison of stimuli $\mathbf{Si}$ at the level of signals $\mathbf{i}$, a reconstruction $S^+(\mathbf{Si})$ of an original signal $\mathbf{i}$ can be used. Recalling that an orthogonal projection from the domain of signals $i \in \mathcal{I}$ onto the subspace $\mathbf{S}$ is achieved by the operator $\mathbf{S}^\top \left(\mathbf{S}\mathbf{S}^\top\right)^{-1} \mathbf{S}$, an optimal implementation for S^+ is $\mathbf{S}^\top \left(\mathbf{S}\mathbf{S}^\top\right)^{-1}$. However, for situations where the inner product $\mathbf{S}\mathbf{S}^\top$ is expected to be close to the identity, this reconstruction can be approximated by the adjoint operator, which in this case corresponds to $\mathbf{S}^\top$. In the context of the tackled problem, the solutions for $\mathbf{S}$ can be expected to allow for the use of the adjoint operator for the purpose of signal reconstruction. The argument supporting this assumption is based on the fact that sensor topologies with positive, non-overlapping receptive fields $\mathbf{s}$ naturally fulfill the constraint $\mathbf{S}\mathbf{S}^\top = \mathbb{D}$, where $\mathbb{D}$ is a diagonal matrix. To assume that $\mathbf{S}\mathbf{S}^\top$ stays close to $\mathbb{D}$ appears plausible considering the following. On the one hand, receptive fields obey $s_j \geq 0$, and on the other hand, although a constraint for non-overlapping visual fields has not been introduced, an excessive overlap of receptive fields would decrease the accuracy of reconstructed signals and can therefore be expected to appear only to a very limited extent. Also note that, in the below proposed formulation, scaling factors contained in the diagonal of $\left(\mathbf{S}\mathbf{S}^\top\right)^{-1}$ can be absorbed by the predictor $\mathbf{P}(\mathbf{M}, \mathbf{q})$. In this way, $\mathbf{S}^\top$ can be used as a reconstruction operator and matrix inversion is avoided, which – from a biological perspective – is appealing since matrix inversion is an operation unlikely to be implemented by neural tissue.

Incorporating the two proposed aspects, an optimization problem for the organization of the given sensorimotor system is written as

$$\begin{aligned} &(\mathbf{S}^*, \mathbf{M}^*, \mathbf{P}^*) = \\ \text{argmin}_{(\mathbf{S},\mathbf{M},\mathbf{P})} \quad & \textstyle\sum_t \left\| \mathbf{S}^\top \left(\sum_k^{n_m} \left({\mathbf{m}_k}^\top \mathbf{q}_t \right) \mathbf{P}_k \right) \mathbf{Si}_t - \mathbf{i}_{t+1} \right\|^2 . \\ \text{s.t.} \quad & \mathbf{S} \geq \mathbf{0},\ \mathbf{M} \geq \mathbf{0},\ \mathbf{P}_k \geq \mathbf{0} \end{aligned} \tag{8}$$

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.

3.4. Method

The organization of $N_s = 16$ visual receptive fields is considered 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. 4 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. 4. The environment is given as a plane textured by a very high resolution image (2448×2448 pixels) depicting a real world scene. The sensor surface is assumed to be parallel to the plane recording grayscale images $\mathbf{i}$ and

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 22 500 triplets $(\mathbf{i}_t, \mathbf{i}_{t+1}, \mathbf{q}_t)$ is obtained via $\mathcal{B}$, where for the presented experiments the underlying policy B selects actions $\mathbf{q}_t$ with sharp activation profiles (all entries in $\mathbf{q}_t$ 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 $(\mathbf{S}^*, \mathbf{M}^*, \mathbf{P}^*)$, the optimization problem given in Eq. (2) is iteratively improved using a projected gradient descent method (Absil et al., 2008). While it is no problem to find a solution with an online method, convergence is much slower, therefore the batch approach is chosen here for practical reasons. However, it is noted 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. 4 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 $\mathbf{P}_k$ were initialized with given random $\mathbf{S}$ and $\mathbf{M}$ to the least squares solution to predict $\mathbf{S}\mathbf{i}_{t+1}$ with $\left[\sum^k \left(\mathbf{m}_k{}^\top \mathbf{q}_t\right) \mathbf{P}_k\right] \mathbf{S}\mathbf{i}_t$ and subsequently projected according to $\mathbf{P}_k \geq \mathbf{0}$. 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.

4. Results

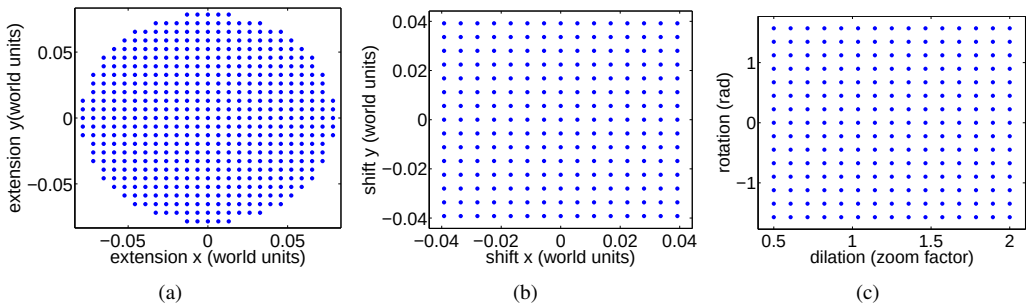
On the basis of the 4-dimensional action space as introduced above, two different sets of sensorimotor experiences, $\mathcal{B}_1$ and $\mathcal{B}_2$, are considered. These two sets of sensorimotor experiences are used to co-develop two sets of sensor and motor topologies $\mathbf{S}_1^*$, $\mathbf{S}_2^*$ and $\mathbf{M}_1^*$, $\mathbf{M}_2^*$. In a first setup, $\mathcal{B}_1$ is recorded using sensor translation actions sampled from a 2-dimensional motor space as shown in Fig. 4(b). Triplets $(\mathbf{i}_t, \mathbf{i}_{t+1}, \mathbf{q}_t)$ in $\mathcal{B}_1$ are sampled choosing actions $\mathbf{q}$ 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 $\mathcal{B}_2$ samples combined sensor rotations and stimulus dilations from a 2-dimensional motor space as shown in Fig. 4(c). As for $\mathcal{B}_1$, triplets $(\mathbf{i}_t, \mathbf{i}_{t+1}, \mathbf{q}_t)$ were sampled with uniform probability from the available discrete actions. Behavior $\mathcal{B}_2$ mimics for example 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. 5 and Fig. 6. They show the optimization process of $\mathbf{S}$ and $\mathbf{M}$ at three different stages. First, the initialization of the problem is shown, then an intermediate stage is shown, and eventually the converged solution is plotted.

The next sections provide a discussion of the optimization process and the emergent properties observed over the course of the optimization. Sect. 4.1 and Sect. 4.2 discuss the formation of spatially compact and Gaussian-like shaped sensory and motor fields from a completely random initialization. Sect. 4.3 discusses the topological organization of sensory and motor fields according to the two different sensorimotor experiences $\mathcal{B}_1$ and $\mathcal{B}_2$.

4.1. Organization of coherent receptive fields

Even though the proposed algorithm is unaware of any topological relationship present in recorded stimuli $\mathbf{i}$, visual receptors cluster as spatially coherent and smoothly overlapping receptive fields of Gaussian-like shape. The reason for this tendency towards spatially coherent receptive fields can be found in the request for an accurate reconstruction of a predicted stimulus $\hat{\mathbf{i}}_{t+1} = \mathbf{S}^\top \mathbf{P}(\mathbf{M}, \mathbf{q}_t)\mathbf{S}\mathbf{i}_t$. Only if $\mathbf{S}$ assumes a form where its rows describe spatially coherent areas (which together cover the entire sensor surface) the original stimulus $\mathbf{i}_{t+1}$ can be reconstructed accurately. For a more visual understanding, one can also imagine to "see $\mathbf{i}$ through $\mathbf{S}$" which results in a more blurred image $\mathbf{S}\mathbf{i}$, the less spatially compact the receptive fields described by $\mathbf{S}$ are.

The development of spatially compact receptive fields over the course of the optimization can be observed in Figures 5(a), 5(c), 5(e), and Figures 6(a), 6(c), 6(e).



(a) (b) (c)

Fig. 4. (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.

4.2. Organization of coherent movement fields

Like for the receptive fields of the sensor, the shape of motor movement fields is not defined a priori. In principle, each movement field could cover the entire or part of the motor space in any manner allowed for by the underlying discretization. The reason for the emergence of spatially coherent and Gaussian-like movement fields is related to the request for an accurate prediction $\hat{\mathbf{Si}}_{t+1} = \sum_k^{n_m} \left(\mathbf{m}_k{}^\top \mathbf{q}_t\right) \mathbf{P}_k \mathbf{Si}_t$, where each $\mathbf{P}_k$ is associated to a movement field $\mathbf{m}_k$, see also Eq. (7). This equation shows that a movement field $\mathbf{m}_k$ defines the area on which the associated predictor $\mathbf{P}_k$ is valid. Consequently, the predictor $\mathbf{P}_k$ is used to predict a future stimulus of an action $\mathbf{q}_t$ for any location $\mathbf{q}$ in the action space where $(\mathbf{m}_k{}^\top \mathbf{q}_t)$ is not 0. And because a linear predictor $\mathbf{P}_k$ can only accurately predict a stimulus for a limited area of the motor space, spatially compact motor movement fields tend to form in order to improve stimulus prediction. In summary: only if motor movement fields $\mathbf{m}_k$ specialize on spatially limited areas, accurate linear combinations of prediction operators $\mathbf{P}_k$ for particular actions $\mathbf{q}_t$ are possible. At the same time, the optimization problem forces movement fields to cover the entire motor area which experiences samples $\mathbf{q}$ in order to be able to assemble predictors $\mathbf{P}^q$ for every $\mathbf{q}$.

The formation of compact movement fields over the course of the optimization can be observed in Figures 5(b), 5(d), 5(f), and Figures 6(b), 6(d), 6(f). In Fig. 6 it can also be observed that movement fields $\mathbf{m}_k$ do not necessarily evolve to be radially symmetric. The reason for this is that – as desired – the shapes of fields $\mathbf{m}_k$ also implicitly depend on the receptive field topology $\mathbf{S}$ in the following way: actions in the motor space which translate the stimulus in directions where the sensor topology $\mathbf{S}$ has a coarser resolution require a less fine-grained resolution of the action space. Thus, along these directions motor movement fields tend to be bigger because the coarser resolution of receptive fields allows a linear predictor $\mathbf{P}_k$ associated to $\mathbf{m}_k$ to achieve accurate predictions over a bigger area of the motor space. This is for example the case for the results obtained for the dilation-rotation sensorimotor experience $\mathcal{B}_2$ shown in Fig. 6 discussed in the next section.

4.3. Overall topological organization

Results presented in Fig. 5 and Fig. 6 demonstrate that with the proposed approach varying sensorimotor experiences induce sensorimotor structures of different macroscopic nature. But before describing these results specifically with respect to sensorimotor experiences $\mathcal{B}_1$ and $\mathcal{B}_2$, the following explanations provide a rationale for how different topologies develop depending on a given sensorimotor experience $\mathcal{B}$.

To better comprehend the organization of obtained sensor topologies (on the left side of Fig. 5 and Fig. 6), the reader is referred back to the work of Clippingdale and Wilson reviewed in the last part of Sect. 1.1 where the fitness of a layout relates directly to the distance between predicted and original point locations. In the problem considered here, just as in Clippingdale & Wilson (1996), a perfect sensor layout is one where receptors exactly map one onto another for every considered action resulting in $\mathbf{P}^q$ matrices where each row contains exactly one non-zero entry. Any deviation from this case leads to an increase in prediction error. Together with insights gained in (Ruesch et al., 2011), this fact allows for the replacement of the Euclidean distance as used by

Clippingdale and Wilson by one based solely on the stimulus prediction error with $\mathbf{P}^q \geq \mathbf{0}$ disregarding any knowledge about the sensor topology.

The obtained layouts for motor movement fields (on the right side of Fig. 5 and Fig. 6), are comprehensible if the previously discussed optimization criteria for the evolution of movement fields $\mathbf{m}_k$ are considered: i) motor movement fields attempt to cover the entire action space which experiences samples $\mathbf{q}$ in order to provide a prediction operator for all those actions, and ii) motor movement fields $\mathbf{m}_k$ tend to be spatially compact to cover a limited area of the motor space in order to evolve an associated predictor $\mathbf{P}_k$ which is specialized for that area of the motor space and which can be used to assemble accurate predictors $\mathbf{P}^q$ via linear combinations using predictors $\mathbf{P}_k$ of neighboring movement fields. As previously described in Sect. 4.2, the shape and size of movement fields $\mathbf{m}_k$ also implicitly depends on the sensor topology $\mathbf{S}$. Shape and size adapt to cover an area of the motor space which allow the associated predictor $\mathbf{P}_k$ of a motor movement field $\mathbf{m}_k$ to provide an accurate prediction of future stimuli. Interestingly, the peak of a motor movement field tends to be located at a position in the motor space which allows for a sensor transformation which maps receptors exactly (or best possible) one onto another. This tendency is related to the fact that at these locations linear predictors $\mathbf{P}_k$ are particularly sparse which, in the case of positive linear predictors $\mathbf{P}_k \geq \mathbf{0}$, also corresponds to situations which allow for a particularly low prediction error. This relationship has been previously described in (Ruesch et al., 2012).

4.4. Topological organization for $\mathcal{B}_1$ and $\mathcal{B}_2$

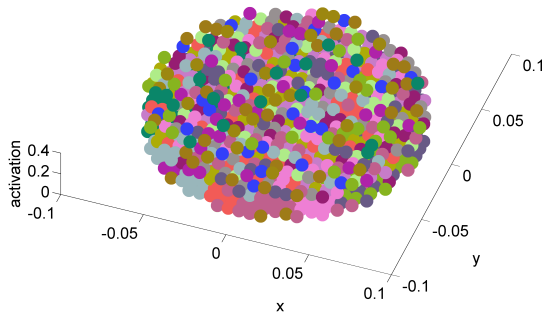
Concluding the analysis of the results shown in Fig. 5 and Fig. 6 in relation to sensorimotor experiences $\mathcal{B}_1$ and $\mathcal{B}_2$, the remainder of this section discusses the characteristics of the obtained topologies depending on the sensorimotor experience used to synthesize them.

In the translation only case (experience $\mathcal{B}_1$), a tendency for hexagonal tiling structures over the entire sensor surface can be identified. In Fig. 5(e) complete hexagonal structures can be observed centered around the blue and pink receptive field. Contrarily, in the rotation and dilation case (experience $\mathcal{B}_2$), receptors organize radially in circular rings. The 3-dimensional perspective of the plots on the left side shows the smooth overlapping between receptive fields. For both cases $\mathcal{B}_1$ and $\mathcal{B}_2$, motor movement fields evolve to cover the translational action space and have a tendency to be placed at locations in the action space which induce sensor translations which map receptors exactly one onto another. For the rotation and dilation case, this leads to elongated movement fields which have a reduced extension in y-direction (rotational actions) and an increased influence area in x-direction (dilation actions). The higher resolution (small extension) of movement fields along the rotational action dimension reflects the higher resolution of receptive fields for rotational actions (9 receptive fields on the outer ring). The low resolution (big extension) of movement fields along the action dimension for dilation actions reflects the low resolution of the sensor in a radial direction (about 2 rings of receptive fields).

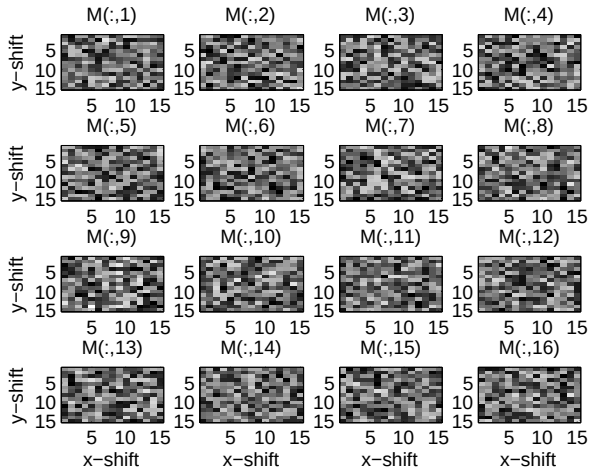
5. Conclusion and outlook

Efficient autonomous robotic systems require a body which is highly adapted to the system's particular task and environment. In biology, a common line of thinking proposes that adaptation implicitly optimizes some underlying criterion which is related to the overall fitness of the organism. However, applying such optimization methods to the design of entire artificial systems is not straight forward. The criteria underlying adaptive processes in biological systems is in general unknown, of considerable complexity, or impossible to evaluate within a reasonable timescale.

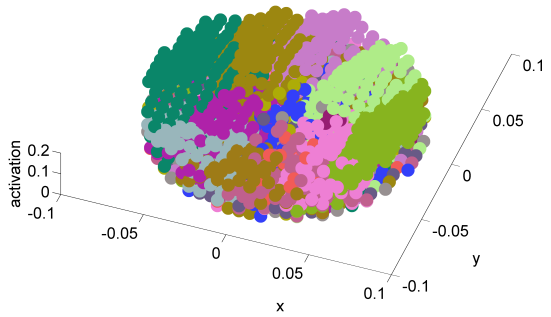
Addressing the development of a computational method for the automated design of behavior-dependent visual sensorimotor structures, this work proposed that it is possible to isolate a simple and at the same time computationally tractable criterion encoding principal characteristics of visual sensorimotor layouts observed in living organisms. We have proposed a methodology that adapts the sensor and motor layouts, as well as stimulus prediction mechanisms, for the agents particular environment and behavioral repertoire. The unified approach for the co-development of these visual sensor and motor structures is based on two main hypotheses. The first proposes that sensorimotor structures can develop according to a general cost function where the agent's behavior and environment are decoupled and enter



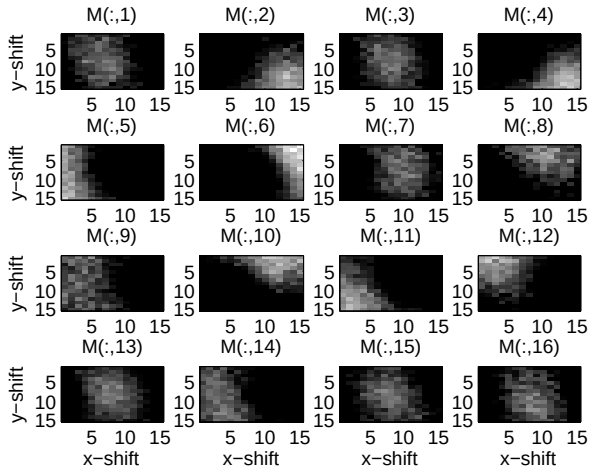
(a) $\mathbf{S}_1$, iteration 1



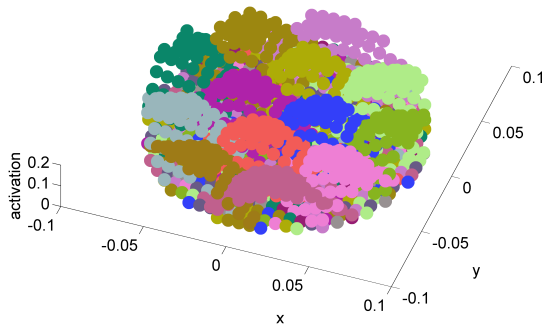
(b) $\mathbf{M}_1$, iteration 1



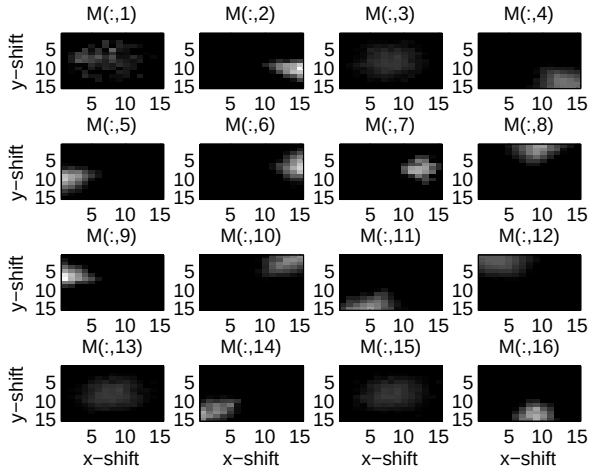
(c) $\mathbf{S}_1$, iteration 80



(d) $\mathbf{M}_1$, iteration 80

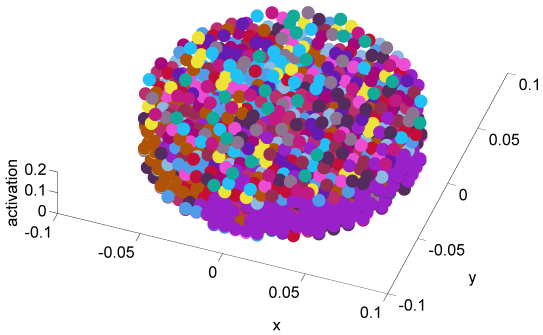


(e) $\mathbf{S}_1$, iteration 400

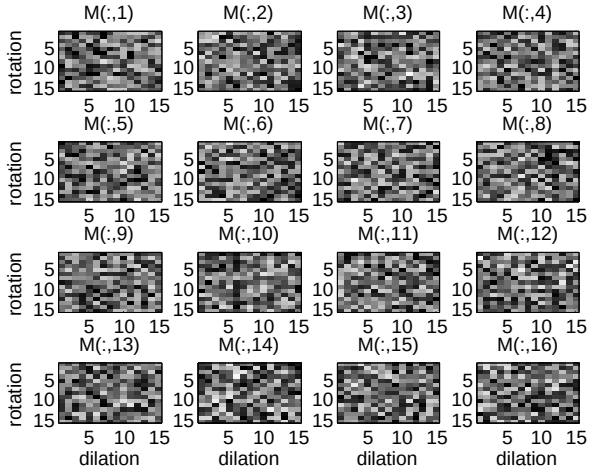


(f) $\mathbf{M}_1$, iteration 400

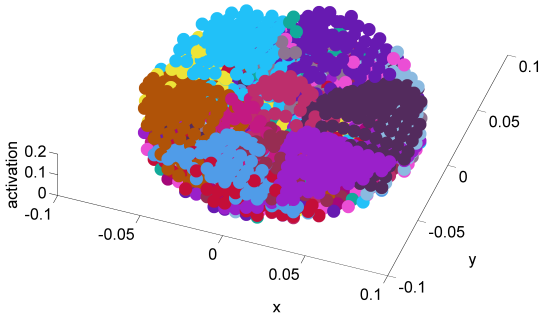
Fig. 5. Sensor and motor topologies obtained for sensorimotor experience $\mathcal{B}_1$. Translation actions were uniformly sampled from a motor space as shown in Fig. 4(b). Left: the evolution of $\mathbf{S}$. Each color denotes a different visual receptive field, and each dot shows the activation of that field at the respective location on the sensor area. Right: the evolution of $\mathbf{M}$. Note, some motor fields happen to overlap, and therefore appear less pronounced as their contribution is combined according to Eq. (7).



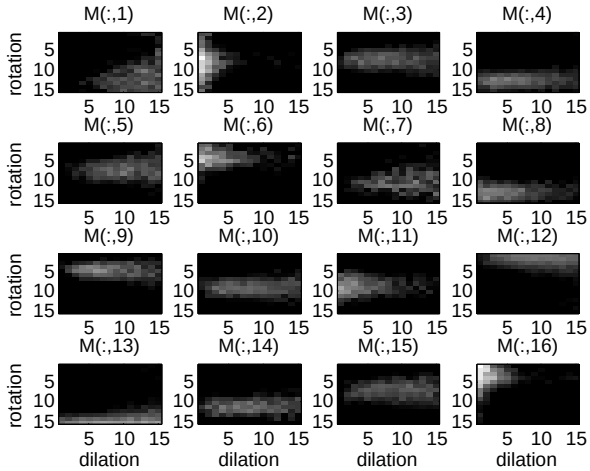
(a) $\mathbf{S}_2$, iteration 1



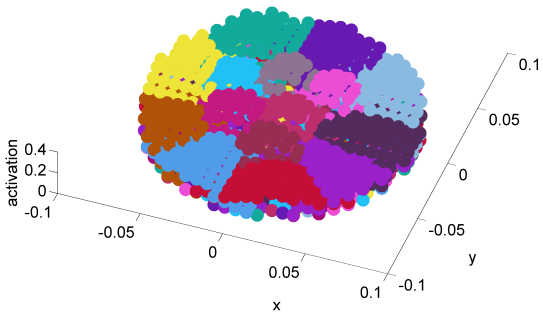
(b) $\mathbf{M}_2$, iteration 1



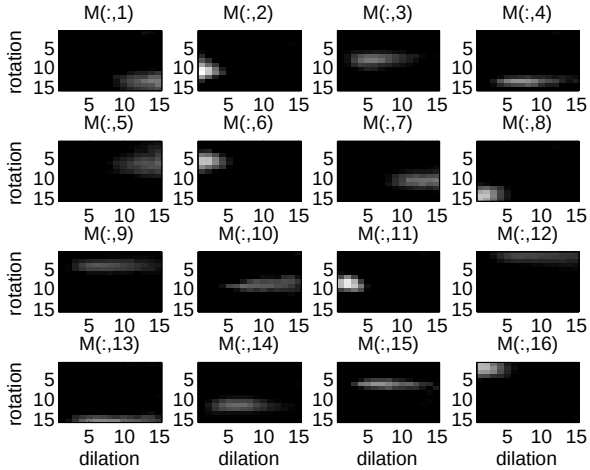
(c) $\mathbf{S}_2$, iteration 80



(d) $\mathbf{M}_2$, iteration 80



(e) $\mathbf{S}_2$, iteration 400



(f) $\mathbf{M}_2$, iteration 400

Fig. 6. Sensor and motor topologies obtained for sensorimotor experience $\mathcal{B}_2$. Rotation and dilation actions were uniformly sampled from a motor space as shown in Fig. 4(c). Left: the evolution of $\mathbf{S}$. Each color denotes a different visual receptive field, and each dot shows the activation of that field at the respective location on the sensor area. Right: the evolution of $\mathbf{M}$. In this case, elongated elliptic fields develop reflecting the higher axial resolution of sensor $\mathbf{S}_2^*$ compared to its radial resolution.

the problem as the agent's overall sensorimotor experience $\mathcal{B}$. The second hypothesis proposes that sensor and motor topologies $\mathbf{S}$ and $\mathbf{M}$ evolve such as to optimize i) the reconstruction of higher dimensional signals, and ii) stimulus predictability.

Per se, it is not clear if the introduced hypotheses are justifiable. However, the proposed framework is capable of reproducing some characteristics of in-nature observed sensorimotor structures, and captures inherent principles present in phylogenetic and or ontogenetic development of biological systems. Therefore, even though the true evolutionary cost function is unknown, it might be claimed that the made assumptions could hold, and that the proposed framework with its simple underlying principles has explanatory power not found in other computational models. Thus, the model can serve to synthesize sensorimotor layouts for artificial visual systems as well as a model capable of explaining some morphological aspects of biological visual sensorimotor systems.

In future work, sensory modalities other than the visual sensory modality could be considered. With the proposed approach this is possible without changes to the principal concept just by replacing the observation and action model. For example, an implementation for an auditory sensorimotor system can be imagined where sensory stimuli and motor actions are recorded and emitted in frequency domain. In this case, receptive elements and motor primitives could specialize on certain frequency bands depending on the interaction of the agent's body and its environment with respect to acoustic properties.

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

P.-A. Absil, et al. (2008). Optimization Algorithms on Matrix Manifolds. Princeton University Press.

C. Cherniak (1995). 'Neural component placement'. Trens Neurosci. 18:522–527.

D. B. Chklovskii (2004). 'Synaptic connectivity and neuronal morphology: viewpoint two sides of the same coin'. Neuron 43:609–617.

S. M. Clippingdale & R. Wilson (1996). 'Self-similar Neural Networks Based on a Kohonen Learning Rule'. Neural Networks 9(5):747–763.

T. B. Crapse & M. A. Sommer (2008). 'Corollary discharge across the animal kingdom'. Nat. Rev. Neuroscience 9:587–600.

D. Floreano, et al. (2004). 'Coevolution of active vision and feature selection'. Biol. Cybern. 90:218–228.

N. Franceschini, et al. (1992). 'From insect vision to robot vision'. Phil. Trans. Biological Sciences 337:283–294.

N. Franceschini, et al. (2007). 'A bio-inspired flying robot sheds light on insect piloting abilities'. Curr. Biol. 17:329–335.

L. Lichtensteiger & P. Eggenberger (1999). 'Evolving the Morphology of a Compound Eye on a Robot'. In Proc. 3rd Europ. Worksh. on Adv. Mobile Robots, pp. 127–134.

J. E. Niven & S. B. Laughlin (2008). 'Review, Energy limitation as a selective pressure on the evolution of sensory systems'. Experimental Biology 211:1792–1804.

B. A. Olshausen & D. J. Field (2004). 'Sparse coding of sensory inputs'. Curr. Opin. Neurobiol. 14:481–487.

G. A. Parker & J. M. Smith (1990). 'Optimality theory in evolutionary biology'. Nature 348:27–33.

C. Paul (2006). 'Morphological computation, a basis for the analysis of morphology and control requirements'. Robotics and Autonomous Systems 54:619–630.

R. Pfeifer & J. Bongard (2006). How the Body Shapes the Way We Think. MIT Press.

R. Pfeifer, et al. (2006). 'Morphological computation for adaptive behavior and cognition'. Int. Congress Series 1291:22–29.

R. Pfeifer & C. Scheier (1999). Understanding Intelligence. MIT Press.

M. B. Reiser & M. H. Dickinson (2003). 'A test bed for insect-inspired robotic control'. Phil. Trans. R. Soc. Lond. A 361:2267–2285.

J. Ruesch, et al. (2011). 'A Measure of Good Motor Actions for Active Visual Perception'. In Proc. Int. Conf. on Development and Learning (ICDL), Frankfurt, Germany. IEEE.

J. Ruesch, et al. (2012). 'Predicting visual stimuli from self-induced actions: an adaptive model of a corollary discharge circuit'. IEEE Transactions on Autonomous Mental Development 4:290 – 304.

N. V. Swindale (2004). 'How different feature spaces may be represented in cortical maps'. Network: Computation in Neural

Systems 15.

R. Tibshirani (1996). 'Regression Shrinkage and Selection via the Lasso'. Royal Statistical Society. Series B 58:267–288.

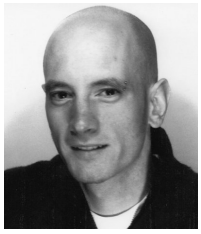
W. E. Vinje & J. L. Gallant (2000). 'Sparse coding and decorrelation in primary visual cortex during natural vision'. Science 287:1273–1276.

B. Webb (2004). 'Neural mechanisms for prediction: Do insects have forward models?'. Trends in Neuroscience 27:278–282.

R. Wehner (1987). "Matched filters'- neural models of the external world'. Comp. Physiol. A 161:511–531.

D. M. Wolpert, et al. (2011). 'Principles of Sensorimotor Learning'. Nature Reviews Neuroscience 12:739–751.

D. M. Wolpert & M. Kawato (1998). 'Multiple paired forward and inverse models for motor control'. Neural Networks 11:1317––1329.



Jonas Ruesch received the M.Sc. degree in Mechanical Engineering in 2005 from the Swiss Federal Institute of Technology, Zurich. In 2013 he earned a Ph.D. degree in Electrical and Computer Engineering from Instituto Superior Técnico, Lisbon. His research focuses on the optimization of adaptive sensorimotor systems, and in particular on the evolution and co-development of sensory and motor structures for active visual perception in embodied artificial agents.

Ricardo Ferreira graduated in Electrical and Computer Engineering in 2004 at Instituto Superior Técnico (IST). In 2006 he received his M.Sc. degree in studying underwater stereo reconstructions of 3D scenes when observed through an air-water interface. His Ph.D. studies focused on reconstructing paperlike surfaces from multiple camera images, and were concluded in 2010. Both his M.Sc. and Ph.D. degrees were obtained at IST. His current research interests focus on general geometric problems arising in robotics and computer vision.

Alexandre Bernardino received the Ph.D. degree in Electrical and Computer Engineering in 2004 from Instituto Superior Técnico (IST), Lisbon. He is an Assistant Professor at IST and Researcher at the Institute for Systems and Robotics (ISR-Lisboa) in the Computer Vision Laboratory (VisLab). He has participated in several national and international research projects and published research papers in top journals and conferences in the field of robotics, vision and cognitive systems. He has been associate editor in large robotics conferences and reviewer for multiple journals and conferences. His main research interests focus on the application of computer vision, cognitive science and control theory to advanced robotic and surveillance systems.