World Scientific
www.worldscientific.com

AUTONOMOUS LEARNING OF THE SEMANTICS OF INTERNAL SENSORY STATES BASED ON MOTOR EXPLORATION

YOONSUCK CHOE*, HUEI-FANG YANG[†]
and DANIEL CHERN-YEOW ENG[‡]

*Department of Computer Science,
Texas A&M University,
College Station, TX 77843, USA*
*choe@tamu.edu
[†]hfyang@cs.tamu.edu
[‡]dce6461@cs.tamu.edu

What is available to developmental programs in autonomous mental development, and what should be learned at the very early stages of mental development? Our observation is that sensory and motor primitives are the most basic components present at the beginning, and what developmental agents need to learn from these resources is what their internal sensory states stand for. In this paper, we investigate the question in the context of a simple biologically motivated visuomotor agent. We observe and acknowledge, as many other researchers do, that action plays a key role in providing content to the sensory state. We propose a simple, yet powerful learning criterion, that of *invariance*, where invariance simply means that the internal state does not change over time. We show that after reinforcement learning based on the invariance criterion, the property of action sequence based on an internal sensory state accurately reflects the property of the stimulus that triggered that internal state. That way, the meaning of the internal sensory state can be firmly grounded on the property of that particular action sequence. We expect the framing of the problem and the proposed solution presented in this paper to help shed new light on autonomous understanding in developmental agents such as humanoid robots.

Keywords: Grounding; semantics; sensorimotor learning; neural encoding/decoding.

1. Introduction

In the emerging field of autonomous mental development (AMD),[1] and in developmental robotics in general,[2] incremental learning is a central issue. In incremental learning, new skills need to be acquired based on acquired skills. One question that arises here is, what forms the minimal basis upon which further learning can bootstrap from? In traditional approaches, domain specific, task specific learning algorithms are used, so new skill acquisition may not be possible at all. In contrast, under AMD (and in developmental robotics), the sensorimotor apparatus

(the robotic body) and a developmental program are all that is needed to form the very basis.[1,3]

However, both of those requirements in themselves pose significant challenges. First, we need to design sensorimotor apparatus that exploit the dynamical properties of the physical world,[4,5] and furthermore, make sure that the capabilities of those apparatus match the ecological task demands[6] (see Ref. 7 for a review). Second, rather than a list of procedural instructions that the agent can merely follow, the developmental program needs to be based on a small number of non-*ad hoc* (and, desirably, biologically grounded) principles which can spontaneously and autonomously give rise to models and goals within the agent, subsequently leading to meaningful behavior.[3]

In this paper, we will focus on problems associated with developmental programing. More specifically, we will look at the very early stage of development (e.g. shortly after birth of infants or at "switch-on" time of AMD robots). We will begin by first considering what is available to developmental programs at that early stage. Then, we will investigate what can (and should) be learned, and what principles should guide such a learning/developmental process.

In order to determine what is minimally available at that early stage of development, we will turn to biology. As mentioned above, agents are born with their sensorimotor apparatus, and that is an absolute minimum requirement. The real question is, what degree of sophistication should these apparatus exhibit? There could be raw sensory arrays and actuators such as photoreceptor arrays and bundles of muscle fibers; or there could be some rudimentary, initial, built-in processes, such as low-level sensory filters and stereotypical motor primitives. In fact, the latter turns out to be the case. Sensorimotor apparatus at birth already contain rudimentary low-level processes, such as center-surround filters in the retinal ganglion cells and the lateral geniculate nucleus, or orientation-sensitive receptive fields in the primary visual cortex (see Refs. 8–10 for reviews); and motor primitives in the central pattern generator (CPG).[11] Note that CPGs can generate these motor patterns in the absence of any sensory input.[12] (Pierce and Kuipers[13] provided some insights on how raw sensors and actuators can be organized into such initial primitives.) Determining how sophisticated the low-level sensorimotor primitives are at birth greatly affects how we should think about developmental programs.

Once we have determined that sensorimotor apparatus and their basic primitives are all that is present at the beginning, it is natural to think of developmental programs being grounded on these existing resources. That is, the first thing developmental programs should do is to learn how to utilize their given sensorimotor apparatus and their primitives to gain understanding of their body and the environment. In fact, research in developmental psychology led to the same conclusion, that development of cognition begins with the sensorimotor stage[14] (also see the related schema theory[15]). The two main questions we will address in this paper are, (i) what can (and should) be learned from the sensorimotor primitives, and (ii) what principles should guide the learning process.

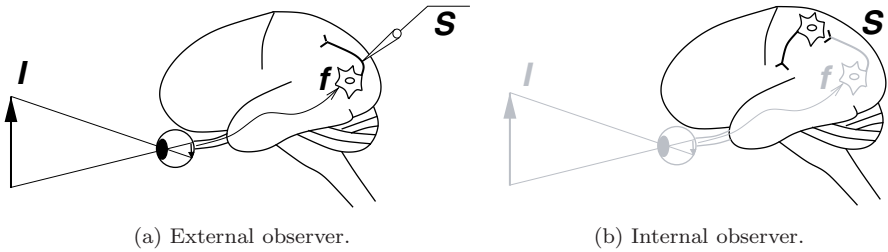(a) External observer. (b) Internal observer.

Fig. 1. External versus internal observer. The problem of decoding internal sensory state is seen from (a) the outside, and (b) from the inside of a perceptual agent. The sensory primitive implemented by a neuron inside the brain performs a transformation from input I to sensory representation s using the function f. The function f is supposed to be a complex one, and deeply embedded in the agent (i.e. it is not at the immediate sensory transduction stage such as the retinal photoreceptors). The task is to find out the property of the input I given the internal state s. (a) With full access to the input I and the state s (as shown as a measuring electrode), what s stands for can be inferred. (b) However, with the lack of that knowledge (the grayed-out area is not known to the second neuron receiving s), such an inference may be impossible. How can the internal observer understand what s means, without peaking outside the skull? (Adapted from Ref. 23.)

Since the sensory input is the main source of information that agents receive (both from their body and from the environment), let us look at what developmental programs can learn, in principle, from these sensory inputs. Let us assume that developmental programs receive input from the sensory primitives, that is, the output from various low-level sensory filters as justified above. One immediate issue with such a setup is that the output from the sensory primitives are *encoded*. For example, we can imagine a visual agent as shown in Fig. 1. The agent receives input I, and its sensory filter f transforms that input into an internal representation s. For example, f could be neurons in the early visual pathway, such as the orientation-tuned neurons in the primary visual cortex, and s the spikes generated by those neurons. The question then is, how can the developmental program know what s stands for? In other words, what is the content of the spikes? The question becomes one of semantics about internal representations/dynamics,[16–19] or one of grounding.[20,21]

For the developmental program to know what s conveys, it either needs to peek outside of the box (as an external observer), look directly at I, and somehow correlate that input to the spike s, as in stimulus reconstruction methods in neuroscience[22] [Fig. 1(a)]. Another way is to know *a priori* the detailed mechanism of the sensory filter f. However, in biology, neither is the case: subsequent stages in the visual pathway receive only the spikes s from the preceding sensory filters [Fig. 1(b)]. The point here is that without any knowledge of the direct input I nor of the low-level sensory primitives f (when output from sensory filters, not raw input, constitutes the first information provided to the developmental program), it is impossible for a developmental program to understand what s represents solely on the basis of s. Without any understanding of its own internal state in such a most basic way, how can autonomous mental development even commence?

One may realize by this point that we have largely left out from our discussion above the other primitive that exists at birth — the motor primitives. In this paper, we will show that in fact these motor primitives are central in providing *behaviorally meaningful* content to the output of the sensory primitives, and thus solve the problem posed in the previous paragraph and in Fig. 1(b). Also, we will propose a simple yet powerful learning criterion for the developmental program — maintained invariance in internal sensory state — which will allow for a meaningful linkage between sensory and motor primitives.

In the next section (Sec. 2), we will provide motivation for our proposed approach and give a sketch of how motor primitives can provide behaviorally meaningful content to sensory primitives. The section thereafter (Sec. 3) will provide details about our learning algorithm, followed by experiments and results (Sec. 4). Finally, we will discuss important issues arising from this work in Sec. 5, and conclude the paper, with a brief outlook (Sec. 6).

2. Learning Behaviorally Relevant Content for Sensory Primitives

In order to see how a developmental program can infer what the internal state of the agent stands for while confined within the agent, let us examine a simple sensorimotor agent with sensory and motor primitives (Fig. 2). The agent has a limited visual field, and the incoming visual signal is transformed via orientation filters (mimicking primary visual cortical neurons) into a spike pattern in the sensory array. As discussed in the introduction, let us assume that the sensory array forms the sensory primitive, thus the agent does not have direct access to the raw input I, and furthermore the property of the filter f is unavailable to the agent. The task of the agent then is to attach behaviorally meaningful content to its own sensory array activity, i.e. to come to an understanding that each sensor represents a certain oriented visual input which can be interacted with in a linear, slanted motion of the visual field.

Imagine we are inside this agent, isolated from the world outside the box, sitting near the π symbol in the middle. It is questionable then whether we will ever be able to associate an orientated visual input stimulus with the spikes generated in the sensor array because we cannot peek outside, and we do not know the particular mechanism of the filters (as discussed in the introduction). Thus, if we view the problem to be the same as Fig. 1(b), there seems to be no solution. However, if we include motor primitives, as in Fig. 2, then there is hope. Why this is the case, we will discuss next.

In Fig. 2, the agent has motor primitives, which basically allows the agent to move its gaze in the direction of the different action vectors. (We discussed earlier that there is reason to believe that such motor primitives exist at birth.) This new addition turns out to be critical. By relating the sensor activity and the motor command that was just generated, *certain aspects* of the sensor property can be inferred. A crucial insight that occurred to us at this point was that certain kinds of
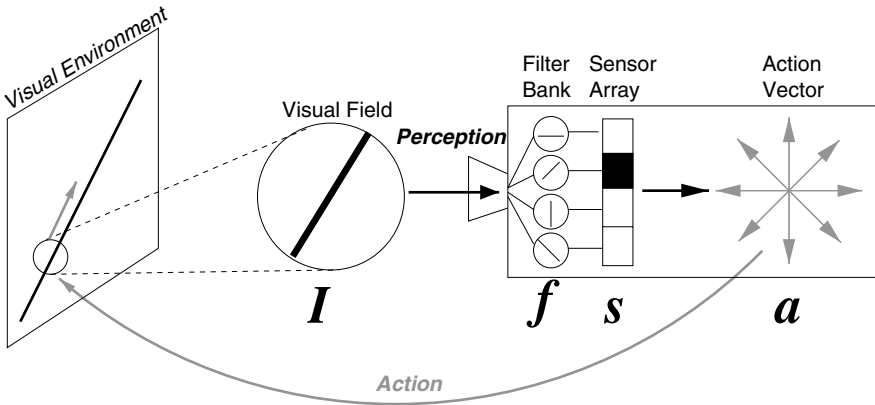
Fig. 2. A sensorimotor agent. An illustration of a simple sensorimotor agent is shown. The agent has a limited visual field where part of the input from the environment (I) is projected. A set of orientation-tuned units f (sensory primitives) receive that input and transform it to generate a pattern of activity in the sensory array s (black marks active). In the example shown here, the $45°$ unit is turned on by the input. Based on the sensory array pattern, a mapping π to motor action vector a is determined, resulting in the movement of the visual field in that direction, and then a new input is projected to the agent. Note that the agent is assumed to be aware of only its internal sensory state s, thus it has no knowledge of I, nor that of f. (Adapted from Ref. 23.)

action tend to keep the sensory activity pattern to remain unchanged (i.e. *invariant*) during on-going movement, and the property of that action exactly reflects the property of the sensory stimulus. Note that by *invariant*, we simply mean that the internal state is unchanging. So, it should not be confused with other uses of the word *invariant*, as in translation, rotation, and scaling invariance in perceptual invariance.

For example, consider the state of the agent as shown in Fig. 3, where a $45°$ input is presented, and the corresponding sensor is activated in the sensory array. Now imagine we move the visual field according to the motor action vector sequence $\nearrow, \nearrow, \nearrow, \swarrow, \swarrow, \swarrow$, which corresponds to a back-and-forth movement along the $45°$ diagonal (along the slanted input). Such an action will keep only the $45°$ sensor turned on during the motor act, i.e. the sensory array will stay invariant over time. We can see that this motion, generated while trying to keep the sensor array unchanged, has led the agent to perform an act, the property of which reflects that of the stimulus. Thus, we are led to conclude that associating sensory spikes with this kind of sensory-invariance driven action itself can potentially serve as the meaning for each sensory neuron. That is, there is a correspondence between (i) the property of the stimulus I triggering a particular sensory state s, and (ii) the property of the action a that maintains invariance in the sensory state s. Thus, a behaviorally meaningful content (the action itself) can be attached to each encoded sensory state, providing a solution to the problem in Fig. 1(b).
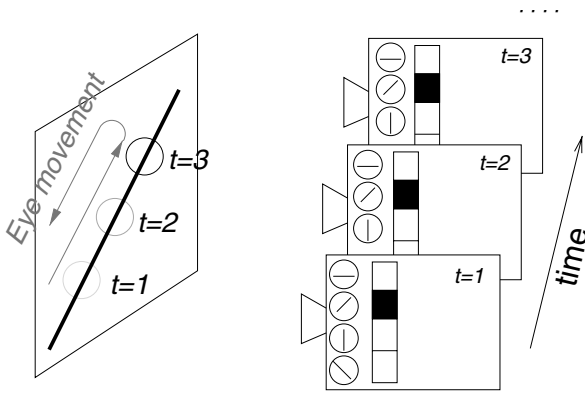
Fig. 3. Invariance in internal state during gaze movement. The internal state of the agent is shown while it is traversing the diagonal input. The internal state does not change over time (i.e. it is invariant) as its gaze moves, and we can observe that the property of the gaze movement (diagonal) is congruent with the property of the input signaled by the internal state. (Adapted from Ref. 23.)

Note that the above scheme can be extended so that invariance is enforced in a spatial pattern (for example, population activity in neurons) or a spatiotemporal pattern, rather than invariance in a single sensory primitive. For example, a sequence of complex actions that give rise to a temporally repeating pattern of activity on the internal state, in which case the two can be correlated. Also, we must clarify that we are not claiming that our approach applies to all sensory modalities. For example, our approach may not work for sensory modalities such as olfaction or taste, where the property of the stimulus (smell) does not easily map to a particular kind of action that also has that kind of property. Certain actions can correlate with stimulus intensity in olfaction or taste, but the associated actions themselves do not possess discriminating characteristics for different odors or tastes. However, our idea can apply well in modalities such as touch. We will discuss more about these points in the discussion section (Sec. 5).

Based on the above insight, we propose that *understanding one's own internal state in terms of one's own action* is one (possibly among several) important principle that developmental programs can abide by, and *internal state invariance* can serve as a simple, yet powerful criterion for implementing the principle. To test the insight developed in this section, we implemented a learning agent based on internal state invariance. The following sections describe the methods and experimental results in detail.

3. Methods

Consider the agent described above (Fig. 2). First, we will describe how input preprocessing was done, and then explain how sensory responses are generated. Next, we define a simple learning rule based on our idea of sensory-invariance driven

action that will allow the agent to learn a mapping from sensory state to *meaningful* action.

3.1. *Input preprocessing*

As a first step, the input image was convolved with a difference-of-Gaussian (DoG) filter[24] to simulate the processing done in the lateral geniculate nucleus in the thalamus. The DoG filter D was defined as:

$$D(x, y) = g_{\sigma/2}(x, y) - g_\sigma(x, y), \tag{1}$$

where

$$g_b(x, y) = \exp\left(-\frac{(x - x_c)^2 + (y - y_c)^2}{b^2}\right) \tag{2}$$

is a Gaussian function with width b, (x, y) the pixel location in the filter, and (x_c, y_c) the center location of the Gaussian peak. For a filter of size $k \times k$, σ was $\frac{k}{4}$ (for all experiments, k was 15), and $x_c = y_c = 8$. The original raw input image I_R (640×480 in size) was convolved with the DoG filter to generate the input I_D:

$$I_D = I_R * D, \tag{3}$$

where "*" is the convolution operator. The input I_D was then subtracted by its pixel-wise mean, and then normalized (divided) by its maximum absolute value so that each element lies within -1 and 1.

$$I_D(x, y) := \frac{I_D(x, y) - \mu_D}{\max_{u,v} |I_D(u, v)|}, \tag{4}$$

where (x, y) is the pixel location, and μ_D the mean of all $I_D(x, y)$. Figure 4 shows an example of I_R and I_D. The final input to the agent (I) was sampled from a 31×31 square area in I_D centered at the agent's gaze location.

3.2. *Internal sensory state and action repertoire*

In order to determine the sensory state, we used oriented Gabor filters. (Gabor filters are known to resemble the receptive fields in the primary visual cortex of mammals.[25]) The Gabor filter G_i is defined as:

$$G_{\theta,\phi,\sigma,\omega}(x, y) = \exp\left(-\frac{x'^2 + y'^2}{\sigma^2}\right) \cos\left(2\pi\omega x' + \phi\right), \tag{5}$$

where θ is the orientation, ϕ the phase, σ the standard deviation (width) of the Gaussian envelope, ω the spatial frequency, (x, y) the location, and x' and y' defined as:

$$x' = x\cos(\theta) + y\sin(\theta), \tag{6}$$
$$y' = -x\sin(\theta) + y\cos(\theta). \tag{7}$$

The formulas above closely follow Ref. 26. Every $m \times m$ filter G_i ($m = 31$) had the same width ($\sigma = \frac{m}{2}$), phase ($\phi = -\frac{\pi}{2}$), and spatial frequency ($\omega = \frac{2}{m}$), while the
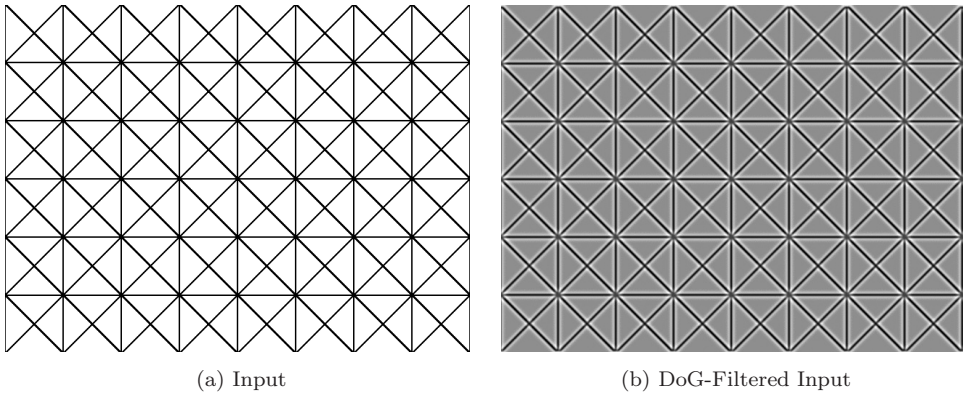
(a) Input   (b) DoG-Filtered Input

Fig. 4. Synthetic input and difference-of-Gaussian-filtered response. A 640×480 synthetic input image consisting of intersecting lines and its Difference-of-Gaussian (DoG) filtered response are shown. A 15×15 kernel was used for the DoG filtering. The DoG response was normalized by subtracting its mean and dividing by the maximum absolute response (black represents negative min and white positive max). See the text for details.

(a) 0°   (b) 22°   (c) 45°   (d) 67°   (e) 90°   (f) 112°   (g) 135°   (h) 157°
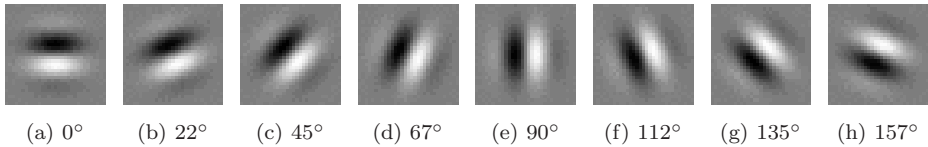
Fig. 5. Oriented gabor filters. The oriented Gabor filters used in the simulations are shown (black represents negative min and white positive max). Each filter was 31×31 in size.

orientation θ varied over the index i ($\theta = \lfloor (i-1)\pi/n \rfloor$, where n is the number of filters). Figure 5 show eight Gabor filters G_i ($i = 1, \ldots, n$, where $n = 8$). (Note that we will use radian and degree interchangeably depending on the context.)

With the definition above, we can now define the orientation filter response. The filter response is a column vector $\mathbf{r}$ where each element r_i ($i = 1, \ldots, n$, where n is the number of filters) corresponds to the sum of the input times the oriented Gabor filter:

$$r_i = \sum_{x,y} G_i(x,y)I(x,y), \tag{8}$$

where (x, y) is the location in the two matrices. (The value r_i is simply the vectorized dot-product of G_i and I.) The vector $\mathbf{r}$ is then normalized with its ℓ^2-norm $||\mathbf{r}||$:

$$\mathbf{r} := \frac{\mathbf{r}}{||\mathbf{r}||}. \tag{9}$$

Fig. 6. Reward probability table. The reward probability function $R(s,a)$ is shown. The rows represent the sensory state (orientation), and the columns motion directions. In this example, four sensory states and eight motion directions are shown. Under ideal conditions, one sensory state would have two best motion directions [thus $R(s,a) = 0.5$ in the two diagonals]. (Adapted from Ref. 23.)

Finally, based on the filter response **r**, a scalar value representing the sensory state s was calculated as:

$$s = \arg\max_{i=1,\ldots,n} r_i, \tag{10}$$

where each value of s corresponds to a unique orientation $\theta = \lfloor (i-1)\pi/n \rfloor$ for $i = 1,\ldots,n$. In this way, given the current input I at the gaze location, s is determined ($s = f(I)$, in the agent: Fig. 2). The set of all possible s values constitutes the set of sensory states S.

For each orientation, there were two matching directions of motion (Fig. 6). For example, for θ of $0°$, the two directions were $0°$ and $180°$. Thus, the set of actions A had $2n$ movement-direction vectors as members (for example, see Fig. 2):

$$A = \left\{ (d\cos(\theta), d\sin(\theta)) \;\middle|\; \theta = \frac{(i-1)\pi}{n}, \text{ for } i = 1,\ldots,2n \right\}, \tag{11}$$

where d is the distance of each movement (fixed to a value of 7), θ the direction of motion, and n the number of orientation filters. Execution of each action vector $(a_x, a_y) \in A$ directs the agent's gaze from the current location (x,y) to $(x + a_x, y + a_y)$.

3.3. *Learning algorithm*

Given a particular sensory state s_t at time t, taking an action a_t takes the agent into sensory state s_{t+1}. The state transition depends on the particular edge feature in the visual scene, thus it is probabilistic. The reward is simply the degree of sensory-invariance achieved across states s_t and s_{t+1}. Let us call this immediate

reward ρ_{t+1}. One way to measure the *degree* of invariance ρ_{t+1} is to calculate the dot-product across the successive sensory filter responses:

$$\rho_{t+1} = \mathbf{r}_t \cdot \mathbf{r}_{t+1}, \tag{12}$$

where $\mathbf{r}_t$ is the filter response vector at time step t [Eq. (8)], and "·" represents the dot-product. With this formula, when the previous filter response is the same as the current, the reward becomes maximized ($\rho_t = 1$) and in the opposite case minimized ($\rho_t = -1$). The benefit of using the vector $\mathbf{r}$ instead of directly comparing the scalar values s is that a graded measure of invariance can be obtained instead of a hard *invariant* or *not-invariant*.

The task of the agent is to learn a state-to-action mapping so that it maximizes the reward ρ_t at time t. However, because of the probabilistic nature of the state transition (which heavily depends on the edge features in the input image), a deterministic state-to-action mapping is not easy to learn. Given the current state s_t, we can think of the conditional probability $P(a_t|s_t)$ so that if we choose action a_t according to this probability, the probability of the next state s_{t+1} being the same as s_t is maximized. Given an estimate of $P(a_t|s_t)$, which we will call the *reward probability function* $R(s_t, a_t)$, we let the agent execute the following policy π at each time step t:

(i) Given the current state $s_t \in S$, randomly pick action $a_t \in A$.
(ii) If a_t equals $\arg\max_{a \in A} R(s_t, a)$,

    (a) then perform action a_t,
    (b) else perform action a_t with probability $R(s_t, a_t)$.

(iii) Repeat steps (i) to (iii) until exactly one action is performed.

To reflect the fact that eye movements follow a fairly straight to slightly curved trajectories between targets,[27] the policy above was augmented with a momentum mechanism where, with a 30% chance, the action from the previous time step $t-1$ was repeated bypassing the steps above. The momentum term's functional role was to aid in the initial stages of training. However, during the final stages of training, the reward probability functions become distinct between preferred and unpreferred directions, thus the momentum term's effect is mostly shadowed by the tendency of moving along the direction signaled by the internal state. Furthermore, the momentum term resulting in a long straight gaze is low: the probability of a straight 6-step gaze is less than 0.14% (= 0.3^6). So, at best, the effect of the momentum is short-lived. In practice, for step ii(b) above, the action was performed if a random draw from $[0, \dots, 1]$ was less than $cR(s_t, a_t)$, where the parameter c controls the strictness of this check (we used $c = 1.55$ during training to relax the check a bit). [Note that, in principle, the entire step (ii) can be replaced by step ii(b): we did test runs with that approach and found that the results were comparable but the simulation was generally slower, especially near the beginning.] When the location of gaze reached the image boundary in I_D, the movement was

wrapped around and continued on the opposite edge of the input. Note that the policy requires the agent to move its gaze every iteration, so a trivial solution of "do nothing" to maintain invariance is avoided. The momentum term also plays a minor role in preventing short, periodic oscillations ($\leftarrow, \rightarrow, \leftarrow, \rightarrow, \cdots$) that trap the gaze in a small local region in the visual environment.

The remaining question is how can $R(s_t, a_t)$ be learned? For that, we used a simple update rule:

$$R_{t+1}(s_t, a_t) = R_t(s_t, a_t) + \alpha\ \rho_{t+1}, \tag{13}$$

where $R_{t+1}(\cdot, \cdot)$ is the reward probability function at time $t + 1$, and α the learning rate parameter ($\alpha = 0.002$). Finally, all $R_{t+1}(s_t, a)$ values were normalized by their sum so that their sum equals 1:

$$R_{t+1}(s_t, a) := \frac{R_{t+1}(s_t, a)}{\sum_{a' \in A} R_{t+1}(s_t, a')}, \quad \text{for all } a. \tag{14}$$

The empirical algorithm above takes elements from reinforcement learning,[28] especially the TD(λ) algorithm (in the extreme case where $\lambda = 1$, i.e. identical to the Widrow–Hoff delta rule).[29] (In fact, a test run with the delta rule $R_{t+1}(s_t, a_t) := R_t(s_t, a_t) + \alpha(\rho_{t+1} - R_{t+1}(s_t, a_t))$, without the divisive normalization, resulted in similar results.) The use of the reward probability function $R(s, a)$ (for $s \in S$ and $a \in A$) is similar to the Q-learning algorithm where the action-value function $Q(s, a)$ is estimated and used to construct a policy.[30] (Also, see our earlier work using Q-learning.[23]) Thus, the main novelty of the algorithm lies in the choice of the reward ρ. The above learning rule converged under all experimental conditions in our simulations (see the next section). One matter of concern here is the use of a small number of quantized states to represent orientation. This concern is related to the fact that we are doing the arg max on the response vector $\mathbf{r}$, rather than using the full information provided by the response vector. This was a compromise we had to take in order to make the learning algorithm simple and efficient. More orientations may be used to alleviate this problem. Furthermore, we are not totally ignoring the full information present in the response vector, since the whole vector is used in measuring invariance ρ, thus the learning rule [Eq. (13)] utilizes the full information.

4. Experiments and Results

In order to test the effectiveness of the invariance criterion in autonomous understanding of internal sensory states, we conducted three sets of experiments. The first was a proof-of-concept experiment with synthetic images with regular grids and diagonal lines as input (Sec. 4.1). The second experiment was with a set of natural images, to show that the criterion is effective in more complex cases (Sec. 4.2). Finally, the learning algorithm was implemented on a pan/tilt web camera, and the results reported in Sec. 4.3.

4.1. *Training and testing on synthetic input*

The first experiment was conducted on a synthetic input with regular straight-line features (Fig. 4). For this experiment, we used four filters ($n = 4$) to match the number of main orientation features in the input image. Consequently, we used eight corresponding motion directions. All other parameters were the same as given in the previous section (Sec. 3). The learning algorithm was executed for 200,000 iterations, during which the reward probability function $R(s, a)$ was gradually updated.

Figure 7 shows the initial, learned, and ideal $R(s, a)$ functions. Initially, for each state s, the $R(s, a)$ values for different a's do not have a particular dominance over others. However, after training, the $R(s, a)$ values are maximized for the direction of motion that exactly *reflect* the orientation of the input that triggered the sensory state s. For example, for state $0°$, the $0°$ and $180°$ directions have the highest $R(s, a)$ value. To emphasize, the important point here is that movement in the horizontal direction based on such an $R(s, a)$ value would have the same property as that of the input giving rise to the sensory state s (i.e. both of them are oriented along the $0°$ axis). The results are close to the "ideal" case shown in the same figure.

For a measure of the performance, we calculated two quantities. First, we calculated the root mean-squared error in the reward probability function $R(s, a)$ compared to its ideal counterpart (Fig. 8). The error value from each iteration is plotted in Fig. 8(a). The error gradually decreases until around 180,000 iterations when it starts to level off. Next, the running average of r_t was calculated and plotted in Fig. 8(b). The running average was calculated as follows:

$$\mu_t = (1 - \alpha)r_t + \alpha\, \mu_{t-1}, \tag{15}$$

where μ_1 was initialized to r_1, and the window-size parameter α set to 0.999. For the average reward, we can see that the slope quickly reaches a plateau around 10,000 iterations. Both measures show empirically that the algorithm reaches convergence.
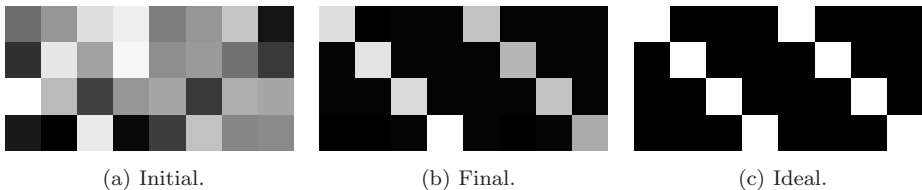

(a) Initial.   (b) Final.   (c) Ideal.

Fig. 7. Reward probability function $R(s, a)$ for each state. The reward probability function values of the four possible sensory states ($0°$, $45°$, $90°$, and $135°$) are shown. See Fig. 6 for plotting conventions (black represents the minimum and white the maximum value). (a) Initially, the rewards are randomly assigned for each direction of motion for each sensory state and normalized so that the sum equals one. The plot shows a "typical" initial condition. (b) After the agent is trained, the reward values become maximal for the movement along the orientations that correspond to that of the input that gives rise to the sensory state. (c) The ideal case shows a diagonal structure, as in Fig. 6.

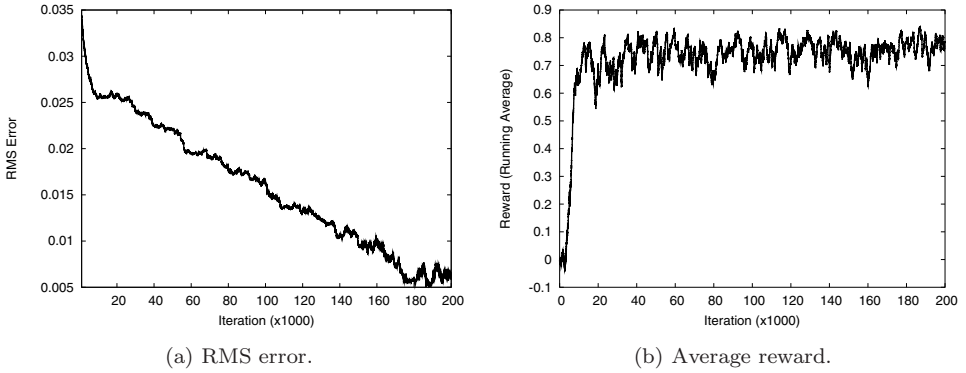(a) RMS error.

(b) Average reward.

Fig. 8. Root mean-squared error in $R(s, a)$ and average reward. (a) The root mean-squared error in $R(s, a)$ and (b) the running average of the immediate reward values are shown. The x-axis represents simulation iteration. The error gradually decreases until it levels off (a), while the average reward rapidly increases to reach a plateau (b).
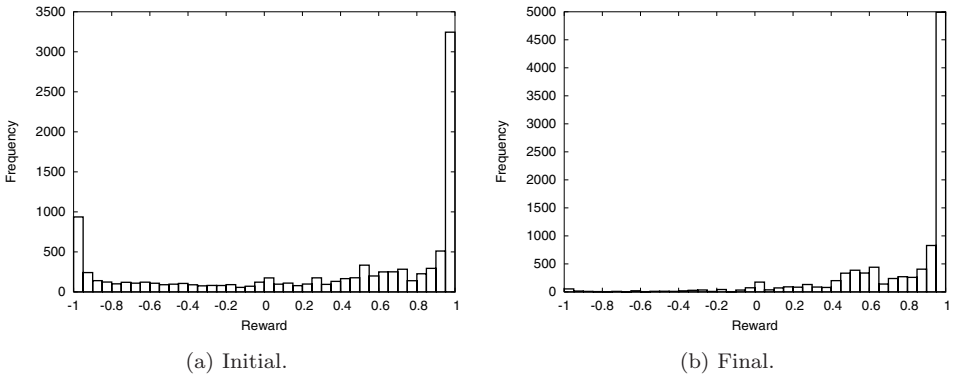

(a) Initial.

(b) Final.

Fig. 9. Initial and final distribution of reward values. The distributions of (a) immediate reward values in the first 10,000 iterations and (b) the last 10,000 iterations are shown. Initially, both positive and negative rewards show a peak (a), but near the end only the positive reward maintains a peak (b).

Since the window-size parameter for the running average was fairly large, it will help to look at the actual distribution of reward values near the beginning and near the end of training. Figure 9 shows the histogram of reward r_t for (a) the first 10,000 and (b) the last 10,000 iterations. Initially, there are two peaks, one at the negative end and the other at the positive end of the x-axis. However, toward the end of training, positive reward dominates. This implies that the policy π (Sec. 3.3) based on the learned probability $R(s, a)$ successfully maintains invariance.

The final test is to observe the gazing behavior of the agent after training. For this test, the learning rate η was set to zero, and the policy π was executed using the learned $R(s, a)$ function. Figure 10 shows example gaze trajectories. As the

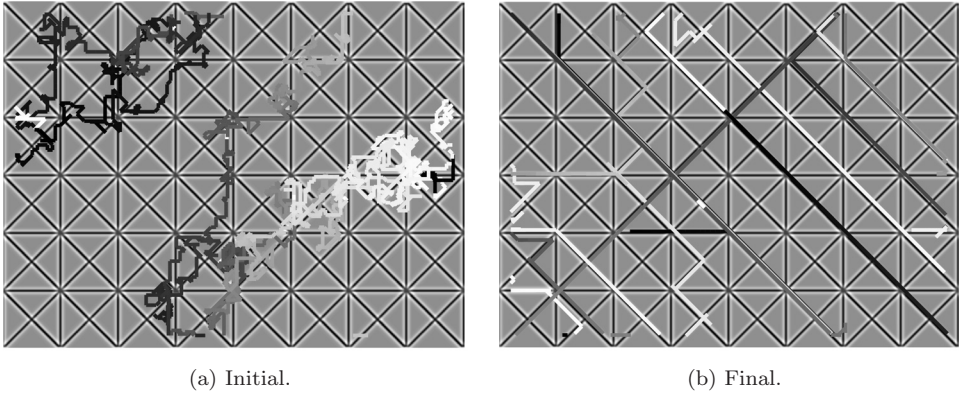(a) Initial.   (b) Final.

Fig. 10. Gaze trajectory before and after training. The gaze trajectories of an agent before and after training are shown for 2,000 time steps each. The trajectory is color-coded so that passage of time is visible (black to gray to white, repeating the whole sequence every 768 steps). (a) Initially, the trajectory resembles a random walk. (b) However, after training, the trajectory shows longer, straighter trajectories that align with the underlying image structure. Note that the important point here is not that the input edges are tracked: What is important is that the resulting gaze behavior is congruent with the sensory property conveyed by the internal state at each moment.

trajectories show, when the agent's internal sensory state is triggered by a particular oriented feature in the input, the agent executes motion whose property exactly reflects that of the current input feature.

4.2. *Training and testing on natural-image input*

Even though the learning algorithm works well under idealistic conditions, it is unknown at this point whether the behavior will generalize to more complex real-world situations. To check if that is the case, we conducted experiments on natural images (Fig. 11). All parameters and experimental setups were the same as before (Sec. 4.1), with the exception that we used 8 sensory filters (and thus 16 motion-direction vectors). The purpose of such a choice was two-fold: (i) to allow a finer-resolution representation of orientation to match the variety of orientations encountered in natural scenes, and (ii) to test whether the algorithm is scalable.

The results are reported in the same format as in the previous section (Sec. 4.1). Figure 11 shows the natural-image inputs used in the training of the agent. Each image was used separately in its own training simulation. The learned $R(s,t)$ function is reported in Fig. 12. Due to the irregularities in edge features in the input and differences across input images, the results are not as clear-cut as before (e.g. state 112° in "Ducks" and a number of errors in state 157°). However, in general, the results are comparable to that in the previous section, and in line with the ideal case [Fig. 12(b)].

The RMS error (Fig. 13) and the average reward (Fig. 14) show similar results as before. There is a small variation in RMS error across inputs, which was expected, but the same convergence behavior is exhibited in all cases. As for the average

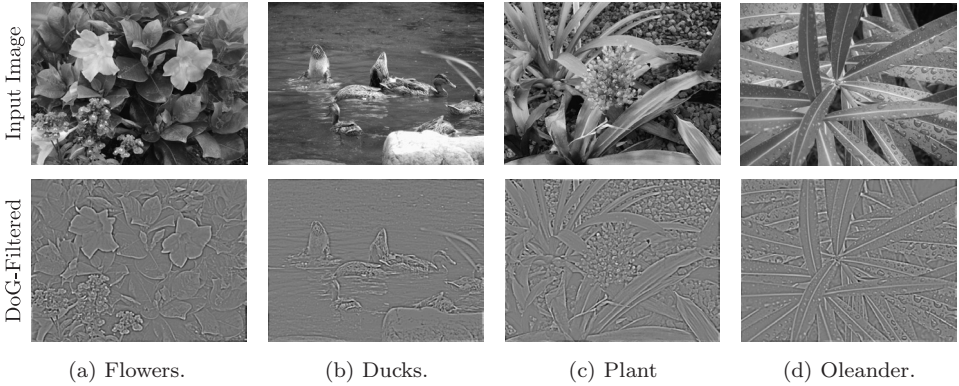(a) Flowers. (b) Ducks. (c) Plant (d) Oleander.

Fig. 11. Natural images and initial preprocessing. Natural images (top row) and their DoG-filtered versions (bottom row) used in the training simulations are shown. Each image was 640×480 in size, and the DoG filter used to produce the bottom row was 15×15 in size. The plotting conventions are the same as Fig. 4. Photo credits: (a)–(c) were taken by Y. Choe, and (d) by James A. Bednar.

(a) Initial. (b) Ideal. (c) Flowers.
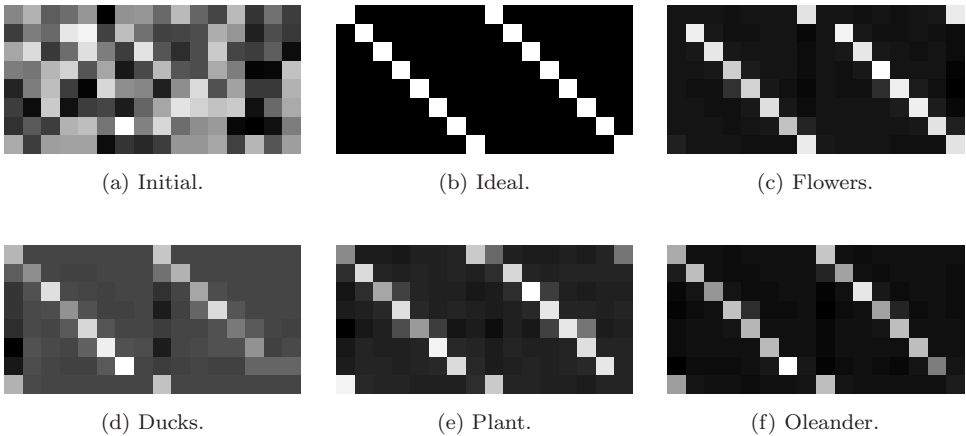
(d) Ducks. (e) Plant. (f) Oleander.

Fig. 12. Reward probability function $R(s, a)$ for natural image inputs. The initial (a), ideal (b), and the learned reward probability function $R(s, a)$ are shown for the 4 natural-image inputs (c)–(f). The plotting conventions are the same as Fig. 7. For the natural images, 8 orientation sensors and 16 motion directions were used, thus 8 rows and 16 columns are shown. The learned motion directions along which $R(s, a)$ values are high closely match the orientation preference of their corresponding sensory state.

reward, convergence to a stable value is slower (again, as expected), and the stable-state reward is lower than that in the synthetic input experiment. Such a result reflects the fact that in natural images invariance cannot be maintained for a very long stretch of time. That is, even though there are many elongated edge features in natural images, they are generally shorter than ideal situations where the edges are very long and straight. The reward distributions (Fig. 15) are also comparable to those from the synthetic input experiment. One distinguishing feature in the

(a) Flowers.

(b) Ducks.

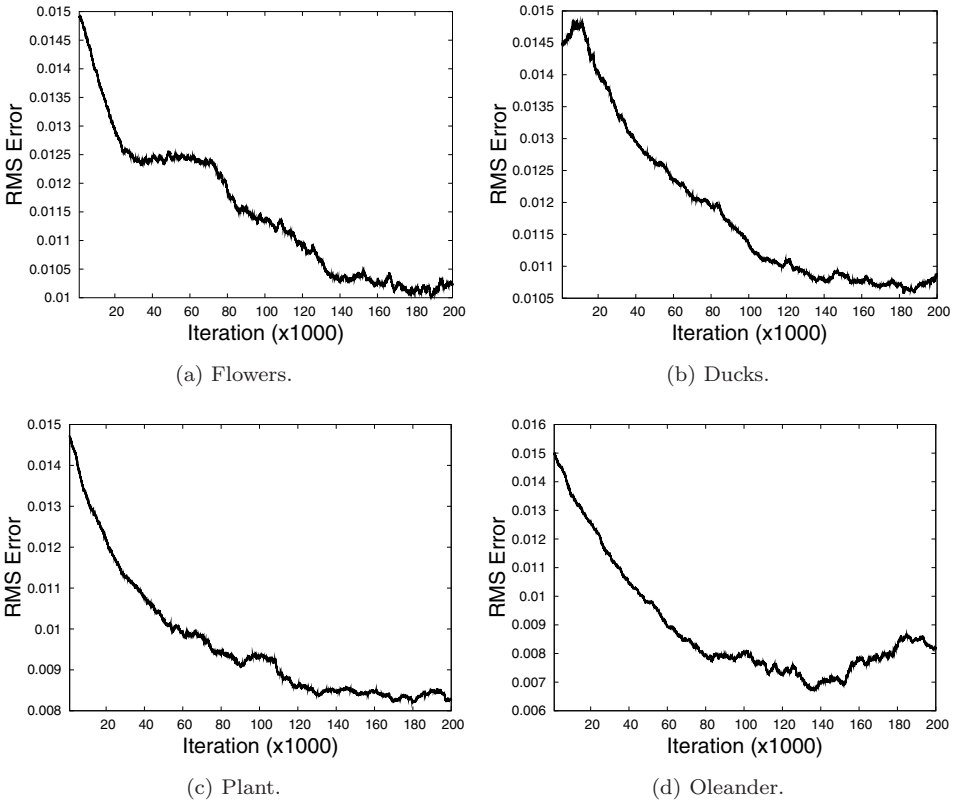

(c) Plant.

(d) Oleander.

Fig. 13. Root Mean-Squared Error in $R(s, a)$ for Natural-Image Experiments. The root mean-squared error in the reward probability function $R(s, a)$ is shown for 4 training sessions using natural-image inputs. As in Fig. 8(a), the error gradually decreases and levels off, but there are small differences across trials.

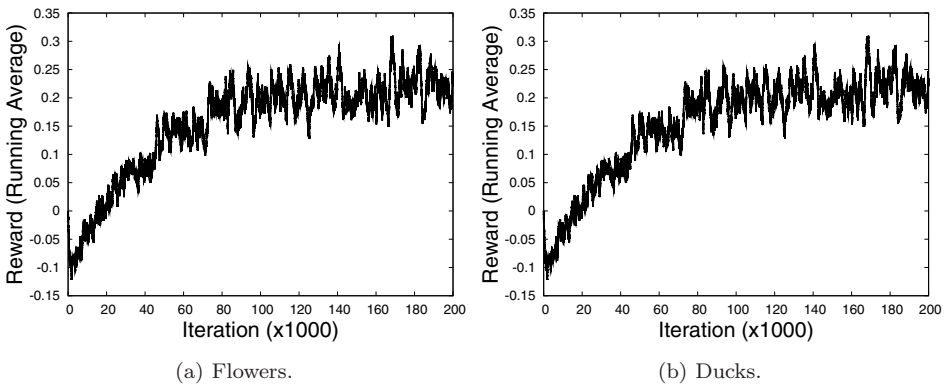

(a) Flowers.

(b) Ducks.

Fig. 14. Running average of immediate reward values for natural-image experiments. The running average of immediate reward values is shown for the 4 natural-image training sessions. Reflecting the fact that straight regular features are less probable in natural images compared to synthetic images, the slope of the increase is slower than that in Fig. 8(b).
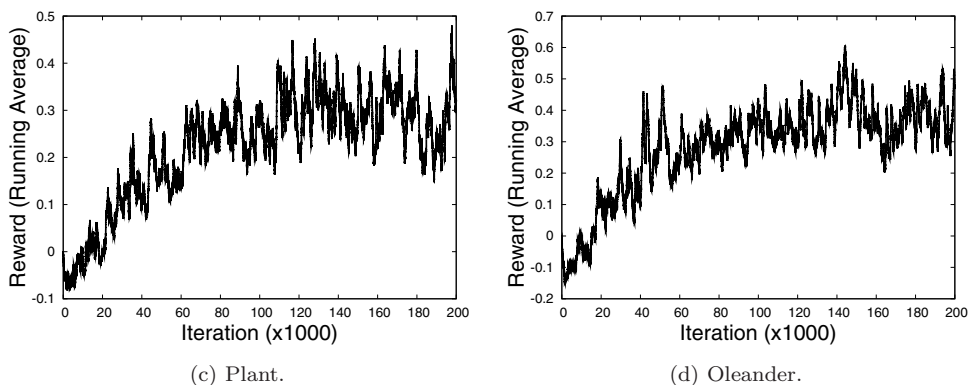
(c) Plant.

(d) Oleander.
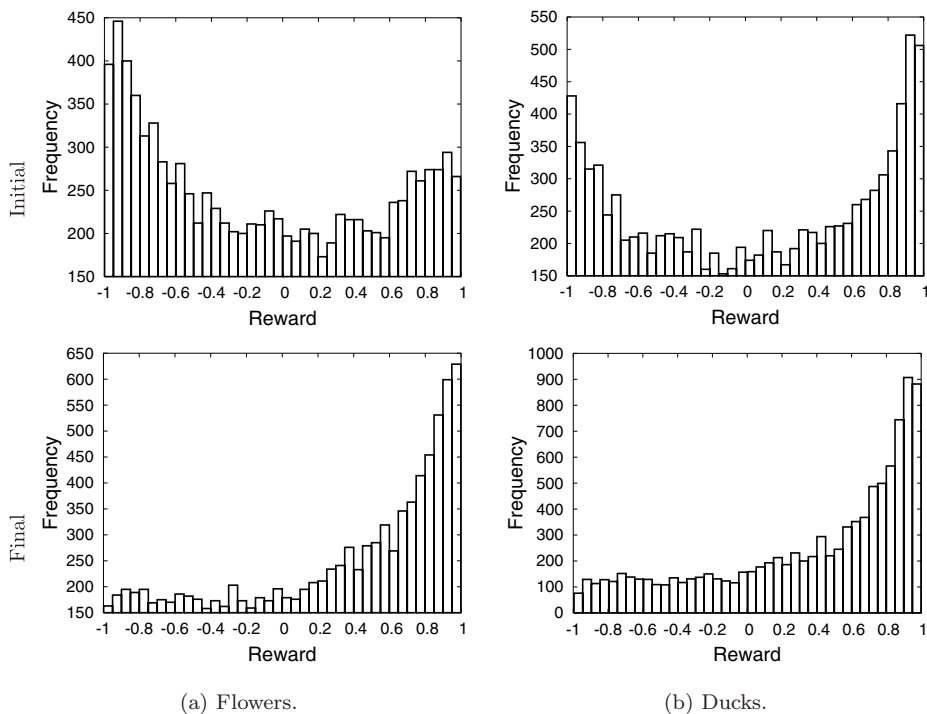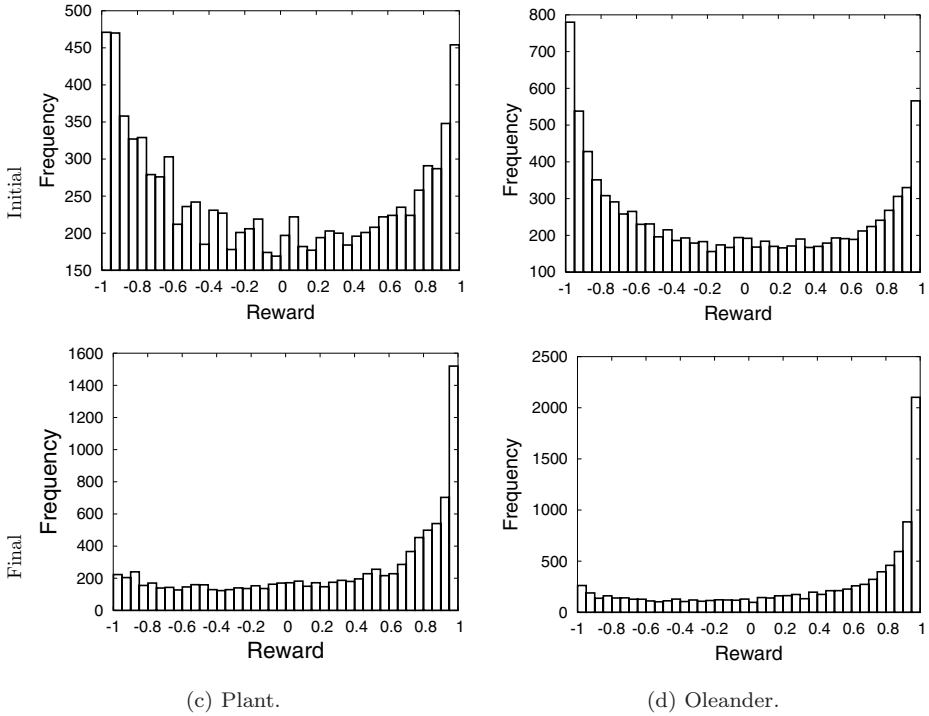
Fig. 14. (*Continued*)

(a) Flowers.

(b) Ducks.

Fig. 15. Initial and final distribution of reward values for natural-image experiments. The distributions of immediate reward values in the first 10,000 iterations (top row, labeled "Initial") and the last 10,000 iterations (bottom row, marked "Final") are shown for the 4 different training inputs. Compared to the initial distribution for the synthetic input (Fig. 9), the peak in the negative reward is more prominent (top row). Consistent with results from the synthetic input, the final distributions have a peak only in the positive reward side (bottom row).

(c) Plant.

(d) Oleander.

Fig. 15. (*Continued*)

histograms is that the distribution is smoother and more gradually changing. Even so, the two initial peaks and the sole final peak in the positive end is the same as before. An interesting observation is that the peak in the final stages of training is sharper for natural image inputs that have more, longer edge features (such as the inputs "Plant" and "Oleander"). The gaze trajectories in Fig. 16 are less well organized than the results in the previous section (Sec. 4.1). This is partly due to the curvy and noisy nature of edges in natural images: the gaze starts off in a certain direction on a curved edge, and as soon as the gaze (which tends to travel straight based on the invariance criterion) goes beyond the point where the gaze is no longer on the edge, then it starts to drift off. Another factor giving rise to this behavior is the fixed step size in which the gaze is shifted ($d = 7$ in Sec. 3). If the step size is greater than the straight local edge then most likely invariance cannot be maintained. On the other hand, if the step size is too small, successive inputs will look almost the same so any action direction will result in invariance, thus rendering the invariance criterion transparent (functionally ineffective). Despite these caveats, on closer inspection, we can see that the behavior in the natural input simulations are *locally* consistent with that of the synthetic input case. For example, if the natural-image-learned $R(s, a)$ values were used to drive the motor policy on the synthetic grid input, behavior similar to that shown in Fig. 10 would result.[31]

(a) Flowers.

(b) Ducks.
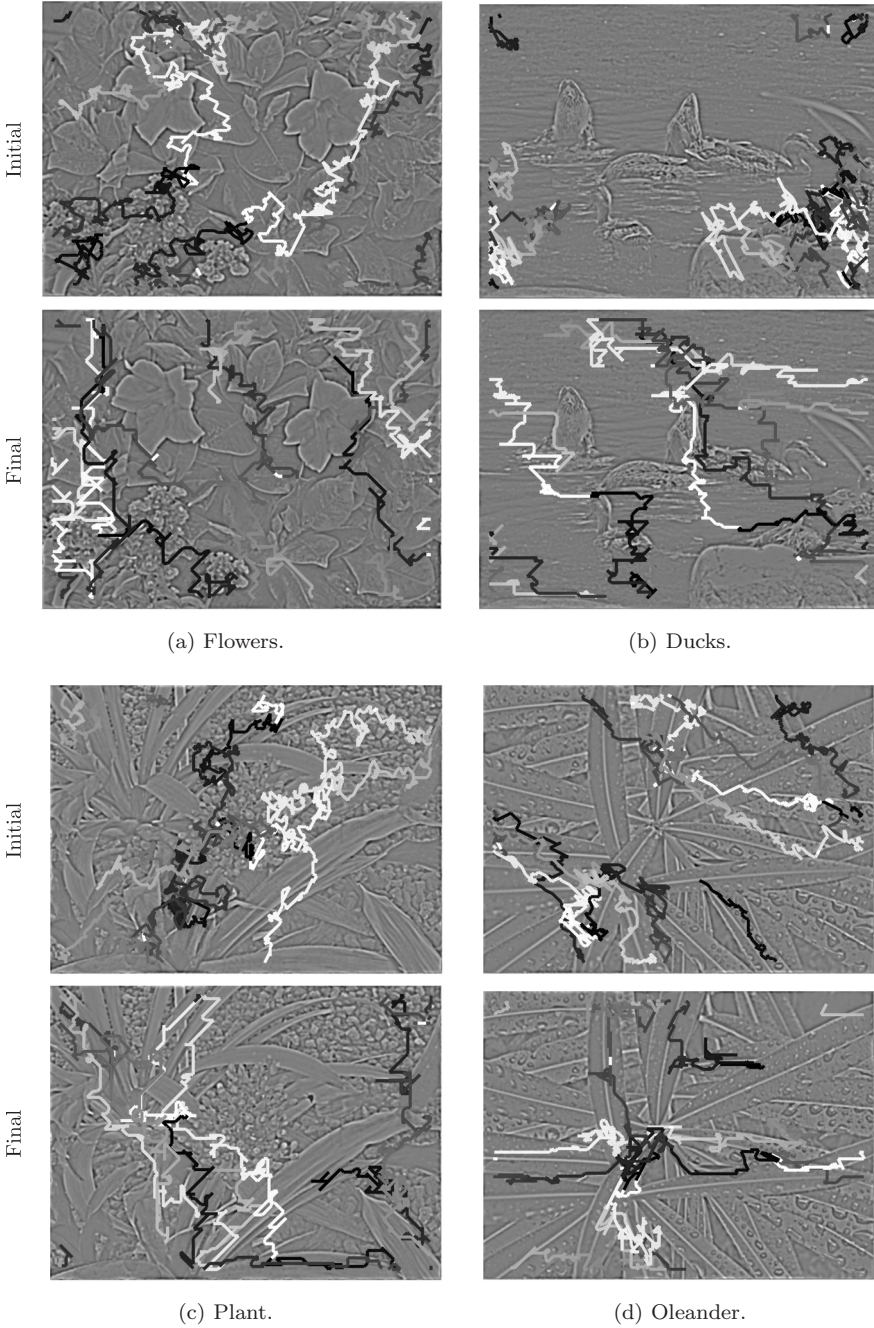

(c) Plant.

(d) Oleander.

Fig. 16. Gaze trajectory before and after training with natural images. The gaze trajectory of an agent trained with natural images is shown for 2,000 steps each. The plotting conventions are the same as that in Fig. 10. The results are comparable to those based on synthetic images (Fig. 10), where the initial trajectories mostly resemble random walk, while final trajectories consist of longer and straighter segments, aligned on the underlying image structure.

In sum, the same invariance-based learning rule allows for the agent to learn to generate behavior whose property closely reflect that represented by its current internal sensory state. The learned behavior subtly varies based on the feature statistics in the natural image, but it maintains an overall uniform qualitative characteristic regardless of the variation in the input.

4.3. *Implementation on a pan/tilt camera*

In order to test the algorithm in a realistic environment, we implemented the learning algorithm on a Pan/Tilt camera (Logitech Orbit Webcam). We used the full $189°$ pan and $102°$ tilt range, and grabbed 320×240 images at 30 frames per second. During this step, the entire image was refreshed based on the current gaze direction. Only a fraction of these captured sequence of images were used, since movement of gaze takes some time and we only needed the images before and after the movement was completed. All training and testing took place on-line.

Figure 17 shows the basic setup (a), (b), and a screen shot of the display window showing the central region as a DoG-filtered gray-scale image (c). The camera was positioned in an office environment, but a house plant was put in front of it to emulate the natural environment. The scene did not contain any moving objects. All simulation parameters were similar to those in the experiments reported above.

The main results are shown in Fig. 18. The learned $R(s, a)$ values show the characteristic diagonal pattern [Fig. 18(a)]. Since the image was refreshed every time the pan/tilt was adjusted, visualizing the gaze trajectory directly is somewhat difficult, so we recorded the pan/tilt angle, translated it back to image coordinates, and plotted it on the initial image took by the web cam. The results are shown in Figs. 18(a) and 18(b). Initial gaze trajectory looks like random walk [Fig. 18(b)], while the final
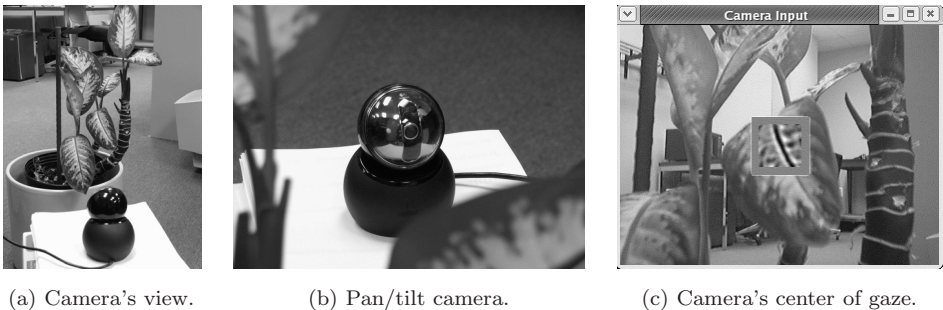

| (a) Camera's view. | (b) Pan/tilt camera. | (c) Camera's center of gaze. |

Fig. 17. Pan/tilt camera. The pan/tilt web camera (Logitech Orbit[tm]) and its operating environment are shown. (a) The camera (the black object near the bottom) and its visual environment is seen from behind the camera. (b) The pan/tilt camera (black) is viewed from its front. (c) A 320×240 window is displaying an image captured by the camera (screen shot). The program shows the center of gaze as a DoG-filtered gray-scale image.

(a) Learned $R(s, a)$.   (b) Initial gaze trajectory.   (c) Final gaze trajectory.
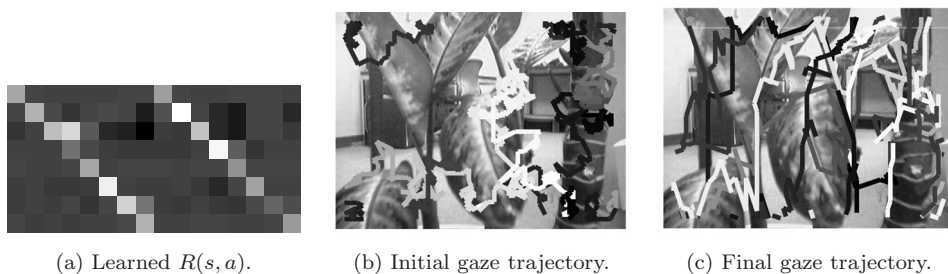
Fig. 18. Results from a pan/tilt camera implementation. The results from running the sensory-invariance driven action learning algorithm are shown. (a) The learned $R(s, a)$ is close to the ideal case, with the strong diagonal components. The (b) initial and the (c) final gaze trajectory are shown. Similar to software-only experiments above, the initial gaze looks like a random walk, while the learned gaze looks more organized, where the action direction provides information about the encoded stimulus property signaled by the internal state.

trajectory shows movement corresponding to the internal state [Fig. 18(c)] at each instant.

4.4. *Summary of results*

In summary, the learning algorithm based on internal state invariance allowed the agent to provide behaviorally meaningful content to its internal sensory state, where the property of the action and the property of the stimulus represented by the sensory state closely aligned. The same algorithm was shown to work very well under diverse input conditions and system configurations. These results support our argument that the invariance criterion can be used to link behaviorally meaningful action with internal sensory states. This way, developmental programs can infer environmental properties from sensory primitives, through action. Furthermore, as the results using the pan/tilt camera shows, the algorithm can be successfully implemented in hardware, and thus can be incorporated into humanoid robot architecture.

5. Discussion

The main contribution of our work is that it showed how a sensorimotor agent can learn behaviorally meaningful action that reflects the property of its internal state, based on an initial set of sensory and motor primitives. Furthermore, a simple, yet powerful criterion of invariance was shown to be effective in linking the agent's internal state with meaningful action. Thus, our work provides insight on what and how developmental programs should learn in the very early stages of mental development: it puts an emphasis on semantic grounding through sensorimotor coordination as one of the first tasks a developmental agent should achieve. In the remainder of this section, we will discuss work that are related to our approach, and various potential issues arising from our work.

5.1. *Related work*

Different Views of Invariance: Invariance, as simple as it seems, has several different meanings. Invariance is an important concept in physics, where a certain measure (such as distance) does not change as the description of a system changes (such as the coordinate frame).[32] In that sense, invariance refers to some kind of property that does not change across transformations or the choice of viewpoints. This is the usual sense of the word when it is applied to perceptual sciences. For example, in object recognition and categorization, properties that are translation, rotation, scaling invariant are sought after.[33,34] Also, there is a tradition in ecological psychology where invariants play an important role. In contrast to the above-mentioned perceptual invariants where an extensive processing may be necessary to extract the invariant properties, the ecological approach asserts that certain invariant properties are readily measurable, directly from the environment[35] (also, see Ref. 6).

In our case, invariance simply means some internal pattern that does not change over time, thus it can be much easier to detect and maintain. This view contrasts with the usual view of invariance outlined above. Both of the approaches above require direct manipulation of the raw sensory input (i.e. invariance exists outside of the agent), unlike our approach based on sensory primitives and internal representations. The theory by O'Regan and Noë is close in spirit to our idea.[36] They pay attention to invariance within the agent (neural state, in their case). For example, they observe that the internal pattern of activity would not change when an agent is gazing back and forth along a line that stretches infinitely. That observation is basically the same as ours in Sec. 2. However, O'Regan and Noë do not make explicit that the property of the back and forth movement exactly reflect that of the straight-line input. They go on to say as much as follows: "...shape in the world can be sensed by the laws obeyed by sensorimotor contingencies...."[36] O'Regan, along with Philipona and colleagues, extended upon the above idea in Refs. 37 and 38. There, they introduce the concept of "compensability." For example, there is an exact correspondence between certain environmental change and certain action that can undo that change. Thus, they are able to assert that a specific environmental-motion/self-motion pair is related through invariance that results from their complementary nature. Such an assertion is getting very close to our own proposal, but there is a subtle difference, because in our case, the internal state maintains invariance even during motion.

In terms of theoretical merits of the invariance criterion, Bell[39] posed a relevant question regarding perception-action cycles. To quote, "What quantity should a perception-action-cycle system maximize, as a feed-forward channel might maximize its capacity?" This is an important question, and we believe our reward criterion of maximizing on-going internal state invariance (or, to be more relevant to the context of Bell's question, maximizing the degree of understanding of own internal sensory state) can serve as one potential answer.

Invariance in the sense that we use it is also known to be effective in coordinating competing motor control laws. For example, Platt *et al.*[40] showed that several competing control laws (which can separately maximize performance criteria in grasping tasks) can be integrated using the concept of "nullspace." They demonstrated that it is possible to execute one control law while change in the performance criteria associated with other control laws remain invariant (the nullspace). However, how such an executed motion and the property of a different invariant performance measure are related in a meaningful way is not made explicit in their paper. For example, Tevatia and Schaal[41] discussed methods to stabilize the end effector, based on an approach that tries to drive the gradient on the nullspace of the Jacobian of a forward kinematic model. This results in the robot stabilizing one of its motor apparatus, so it conflicts with our algorithm's requirement for patterned motion. On the other hand, their work suggests that invariance as unchanging internal pattern may have a broader applicability than what we proposed in this paper.

Lastly, we wish to emphasize that invariance *per se* should not be equated with content or meaning of the internal sensory state. Invariance should be viewed as a simple enough criterion to follow which will eventually lead to the linkage of internal sensory state and behaviorally meaningful action, and that very action is what constitutes the content or meaning. Invariance itself is not a goal but a criterion. Rather, maximization of *understanding* is the goal (or principle).

The Role of Action in Sensory Understanding: There are many previous works that recognize the importance of voluntary action in sensory understanding. These works help strengthen our argument that behaviorally meaningful action provides content for internal sensory states. For example, Held and Hein[42] showed that animals reared to receive identical visual input greatly differed in their development of visual capabilities depending on whether the animal was passively receiving input or actively interacting with the environment. Another good evidence that action provides perceptual content is the tactile-vision sensory substitution (TVSS) experiment by Bach-y-Rita and colleagues.[43,44] In their experiments, blind subjects were able to partially gain visual function through a tactile array which conveyed signal from a video camera. One crucial component in making that work was voluntary action. Under passive viewing conditions, vision-like perception did not occur in the subjects. Applying our analysis to the two examples above, we can explain why this should be the case. Passive viewing only stimulates the sensory primitives, thus no content can be associated to it. However, once one starts to move, the sensory primitives can be linked to a particular mode of action, through the invariance criterion.

The field of behavior-based robotics by Brooks and colleagues is also quite relevant in this context.[5,45,46] The robotic architecture in these works build up from a reactive to a progressively contemplative "layers" in control architecture. At this stage, our work is mostly at the reactive stage (no memory), but with some amount of memory, more complex behavior-based semantics can be learned (see Sec. 5.2 for a discussion on this point).

There are also notable efforts to identify different principles underlying learning and development in robotics. Edsinger and Kemp[47] showed how the concept of controllability can be used to pick perceptual categories that are more relevant to the robot's actions. Contingency has also been suggested as a possible principle to guide learning, as in the work of Gold and Scassellati.[48] The basic idea is to distinguish between sensory signal caused by the robot itself versus those by other agents in the environment by analyzing the delay between action generation and sensory feedback.

Several other efforts are more closely linked to the work we presented here. Cohen and Beal proposed an approach similar to ours in spirit where they put emphasis on action in developing "natural semantics."[49] Their work is also inspired by questions in infant development: what is initially available and what should be learned. Their concept of "image schemas," borrowed from Mandler (Ref. 50, as cited in Ref. 49) is in fact very close to the idea that action should be integrated into the content of sensory states. However, they do not provide an explicit link between the properties of sensory input and action, nor a simple, biologically implementable criterion (such as invariance). Pierce and Kuipers also looked into developmental issues, but they went deeper than us and asked what if the sensorimotor primitives did not exist and the agent only had raw sensors and actuators.[51,52] They used statistical methods to form the sensory and motor primitives, but again, they did not make an explicit link between the two. (The work by Bowling *et al.* on "Action Respecting Embedding" shows how the motor system can constrain sensory primitive development.[53]) However, from their work, we can gain insight on how sensorimotor primitives themselves can evolve or develop. The work by Singh *et al.*[54] may be a good direction to expand our work with the layered architecture in Brooks *et al.* Singh *et al.*[54] showed how intrinsic reward (such as saliency) can be used along with environmental cues to construct a hierarchy of action skills. Note that in this work, they emphasized the intrinsicness just as we did, while they used a different criterion, that of salience, rather than invariance. This way, their learning agent can relate its internal states signaling salient events with action that can directly trigger those salient events. So, in this sense, those triggering actions somehow provide meaning to the internal states.

Finally, the work on self-organization of primary sensory areas, such as the visual cortex, provides similar insights on sensory primitive formation,[10,55] and early results show that these approaches can be integrated with the approach presented in this paper[56] (Fig. 19). This is a potentially productive future direction, and we are actively working on related issues.

5.2. *Potential issues*

Dealing with Complex Internal States: The model presented here is decidedly simple to convey the essence of the problem at hand, and as such, it can be extended in several directions. We would like to note that internal-state invariance does not

(a) Learned receptive fields.
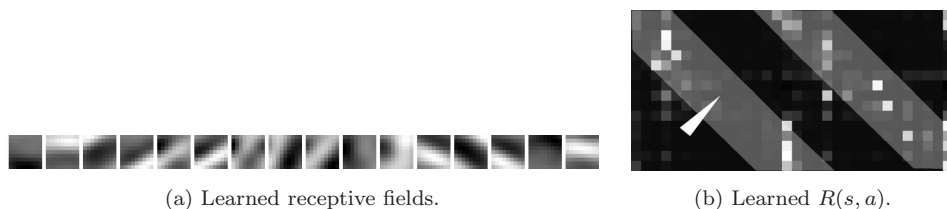

(b) Learned $R(s, a)$.

Fig. 19. Results from co-development of receptive fields and $R(s, a)$. Preliminary results from co-development of visual receptive fields and the motor policy ($R(s, a)$) are shown. The receptive fields were ordered so that they show $0°$ to $180°$ orientation at a fixed interval. The rows in $R(s, a)$ were reordered in the same manner so that an interpretation identical to Fig. 6 is possible. The receptive fields were trained based on a combined Hebbian/invariance criterion, and $R(s, a)$ was trained as usual. The receptive fields developed into Gabor-like orientation-tuned shapes, and the reward table $R(s, a)$, although not perfect, shows the characteristic diagonal component. This is more pronounced on the right half shaded gray than the left half shaded gray (the area marked with the arrow is not strongly represented). These are encouraging results, but more work needs to be done, which we leave for future work. (Adapted from Choe and Yang.[56])

always have to be defined on a single neuron's activity. Any kind of pattern, be that spatial or temporal, can be attempted to be maintained invariant while performing an action. Thus, meaning based on action can also be ascribed to a repeating pattern of activity, not just to a single spike. Also, invariance can be maintained in only one part of the global pattern, which can be seen as a relaxed version of the invariance criterion. Attentional mechanisms[57] may be necessary for this purpose. We believe investigating in this direction will be most fruitful, and in fact we are currently steering our effort into this problem (see e.g. Ref. 58).

Action as Content: One criticism against our approach can be that our model implicitly assumes that knowledge of one's own action is immediate (see Ref. 59, pp. 60–61, for some insights on this issue). Such an attack does not take into account that all we are claiming is that *behaviorally meaningful action* gets associated with a particular sensory state. What is important is that that action is behaviorally meaningful, because the property of that action and the property of the corresponding sensory state closely match. Also, such an action allows the agent to interact with the sensory input in a meaningful way. In that sense, that action *directly* serves as content, or meaning for the sensory state. Thus, it is not necessary to possess a meta-level knowledge about the action itself; "you just need to know how to act."[60] Understanding of something, in the most primitive sense, can be defined as acting out in a manner which is relevant to that something which is understood, since possessing some kind of knowledge is not useful unless it is *eventually* expressed in some form of behavior. Some may feel that all the above is too behavioristic, depending too much on automatic reflexes. The simulations of our model are in fact largely reflexive, but they do not need to be. More cognitive (or contemplative) operations can be grounded on the same principle we advocated in this paper. What makes behavior less immediate depends on complex issues such as emergence

of goals, attention, motivation, etc. within agents (see Ref. 61). However, discussing those would go far beyond our initial goal of investigating the very early stages of mental development, thus we will leave that topic as future work.

Applicability to Other Sensory Modalities: Can our approach be extended into other sensory modalities such as audition, touch, olfaction, etc.? Our approach is general enough to be easily extended into certain modalities such as touch (see e.g. Ref. 62), but it cannot work very well in domains where there is not much correlation between action and the perceived sensory state, e.g. olfaction. There is however some evidence that the act of sniffing can alter the perceived sense of smell,[63] which indicates that our approach may also be applicable, although in a limited way, in the olfactory domain. Another domain where our approach cannot be effective is color vision. The reason is that no patterned action correlate with shades of color in a lawful way. On the other hand, there are interesting experimental results that show artificially altering the sensorimotor law regarding color can change the subjective experience of color.[64] Whether sensory and motor primitives can be found that can be matched to maintain invariance in color is an interesting open question.

Here, it would be useful to mention a different kind of content, those that are related to reinforcement signals such as gustatory rewards. For example, Rolls proposed in Ref. 65 that semantics of brain states can be either related to (i) sensorimotor skills (as we also proposed) or to (ii) environmental reinforcement. Olfaction, gustation, and other such sensory modalities convey signals that are highly related to survival values, and thus they may take on the latter kind of meaning.

Applicability to Different Levels and Configurations of Internal State: Another criticism we anticipate is that if the agent had a rich array of sensors, such as a 2D matrix of RGB pixels, then the properties of the visual environment can be easily recovered within the agent through unsupervised learning even without direct access to the outside world. However, this does not help solve the problem, because such a rich information is only available at the very first stage of sensory processing. The next stage, and the stages following that, etc. only receive a more and more *encoded* version (the output from the unsupervised learning algorithm) from the previous stage, just like the sensory array in our agent which receives only encoded spikes from the orientation-tuned filters. In artificial agents, a detailed photographic input will also lead to problems, because in that case, the problem of visual understanding becomes too complex. In fact, the problem faced by the agent becomes the same as that of computer vision, which is not a trivial one by any means.

Another related criticism may be that the learning criterion we proposed may not be applicable to all stages of sensory processing. For example, the retinal ganglion cells and the lateral geniculate nucleus (LGN) in the early visual pathway show a center-surround receptive field property. It is not easy to imagine any kind of action sequence that would possibly keep these neurons' activities invariant. Our response to this is that sensorimotor coupling may not exist at such an early stage of sensory processing (see e.g. Ref. 66), and thus we do not claim that our approach will work

in this stage. A corollary to this observation is that at these low levels, invariance anchored on action may not be observed.

On a similar note, one can argue that meaningful stimulus properties can be extracted from a population of signals even when they are highly encoded. For example, instead of looking at a single orientation filter output, what if we got many such outputs? Could it be possible to infer the stimulus properties that gave rise to such a pattern of activity, solely on the basis of that sensory filter response? For example, co-occurrence statistics of internal state activity can be analyzed to infer stimulus properties. In our view, such an inference will require unnecessary prior knowledge, such as: (i) the orientation filters are laid out in a topology-preserving manner that reflects the spatial structure of the input space, and (ii) filter activity aligned along a straight line in the internal state space indicates that those individual filters represent collinear orientations in the input space. Thus, such an approach cannot work under conditions allowed in our agent model.

Using different kinds of encoding, such as firing rate or interspike interval instead of spikes, does not help either, again if the analysis is to be based on the sensory filter response alone. That is because in whatever form they are, the signals convey encoded sensory information. Thus the problem associated with decoding remains the same.

Biological Plausibility and Predictions: In this paper, we mostly dwelled in a rather abstract level when discussing developmental programs and the principles they should go by, even though we talked about visual receptive fields and often times used the metaphor of neurons. For example, we did not mention at all where the motor primitives may reside, nor at what level in brain organization does invariance get enforced. In our view, there is no single place or level where invariance is maintained. According to Fuster, there is potentially sensorimotor linkage at all levels of brain organization, except for the very low levels such as immediate sensory surfaces[66] (see Ref. 45 for similar views in robotics). Thus, it is reasonable to assume that the invariance criterion can be employed at multiple levels of sensorimotor organization.

One example is the recently discovered "mirror neurons" where the property of a particular articulated action that trigger activity in the neuron almost exactly aligns with visually perceived gesture that activates the same neuron.[67,68] An intriguing point here is that such a sensorimotor linkage happens at a single-neuron level. From a developmental point of view, it may be interesting to see how the action–gesture linkage develops in the mirror system. Our prediction is that the invariance criterion may play an important role in this system.[69] The first thing to test is whether the mirror property exists at birth. If it does it would refute our idea, since we predict that the mirror property develops over time after birth using the invariance criterion.

Testing whether the invariance criterion is in use is a much more trickier issue. One way to indirectly test this idea is with the tactile sensory substitution system by Bach-y-Rita and colleagues.[70] In their most recent setup, they used electrotactile

arrays on the tongue to convey diverse sensory modalities, such as vestibular information. Two different electrotactile representations can be provided to the subject. In the first case the property of the encoded signal is to be easily correlated to particular mode of action so that it is easy to maintain invariance in the electrotactile input. For the control case, the electrotactile representation is to be generated based on some other principle, such as information maximization. Our model predicts that performance will be much higher in the former case. Another alternative is to mix different representations based on different principles in the same electrotactile representation, and measure which part of the mixed representation gets utilized. Again, the prediction is that the part of the representation that is amenable to the invariance criterion will get selective attention.

Finally, we can ask what is the biological basis of the internal-state invariance criterion itself (i.e. the reward used in the learning algorithm). One possibility is to relate the criterion to the concept of "homeostasis," much developed in cybernetics and systems theory (see Ref. 71 for a review). Homeostasis is widely observed in organisms and organismal function, and its evolutionary value is clear, because homeostasis gives an organism persistence over time. Interestingly, these processes also seem to exist in the nervous system, and can be observed at the level of synapses as shown by Turrigiano and colleagues[72,73] (see Sullivan and de Sa[74] for a computational model). Although these homeostatic properties do not exactly map to the internal-state invariance in our framework, it is notable that the general property of persistence is prevalent in biology, including the neural system.

6. Conclusion

In this paper, we analyzed potential problems associated with the very early stages in autonomous mental development (AMD), and laid a foundation upon which subsequent stages of development can bootstrap from. We proposed *understanding of own internal sensory state* as a principle to be employed by AMD, and identified *internal-state invariance* as an easy-to-implement criterion that can be used toward that end. We expect our approach, in the long run, to provide new insights on how agents such as humanoid robots can acquire "natural semantics" in an autonomous manner, in terms of their own sensorimotor apparatus.

Acknowledgments

References

1. J. Weng, J. L. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur and E. Thelen, Autonomous mental development by robots and animals, *Science* **291**(5504) (2001) 599–600.

2. M. Lugarella, G. Metta, R. Pfeifer and G. Sandini, Developmental robotics: A survey, *Connect. Sci.* **15** (2003) 151–190.

3. J. Weng, Developmental robotics: Theory and experiments, *Int. J. Humanoid Robot.* **1** (2004) 199–236.

4. R. D. Beer, Dynamical approaches to cognitive science, *Trends Cogn. Sci.* **4** (2000) 91–99.

5. R. A. Brooks, Intelligence without representation, *Artif. Intell.* **47** (1991) 139–159.

6. R. Pfeifer and C. Scheier, *Understanding Intelligence* (MIT Press, Cambridge, MA, 1999).

7. O. Sporns, Embodied cognition, in *The Handbook of Brain Theory and Neural Networks*, ed. M. A. Arbib, 2nd edn. (MIT Press, Cambridge, MA, 2003), pp. 395–398.

8. N. Daw, *Visual Development* (Plenum Press, New York, 1995).

9. E. R. Kandel, J. H. Schwartz and T. M. Jessell, *Principles of Neural Science*, 4th edn. (McGraw-Hill, New York, 2000).

10. R. Miikkulainen, J. A. Bednar, Y. Choe and J. Sirosh, *Computational Maps in the Visual Cortex* (Springer, Berlin, 2005), [http://www.computationalmaps.org].

11. E. Marder and R. L. Calabrese, Principles of rhythmic motor pattern production, *Physiol. Rev.* **76** (1996) 687–717.

12. R. Yuste, J. N. MacLean, J. Smith and A. Lansner, The cortex as a central pattern generator, *Nature Rev. Neurosci.* **2** (2005) 477–483.

13. D. M. Pierce and B. J. Kuipers, Map learning with uninterpreted sensors and effectors, *Artif. Intell.* **92** (1997) 162–227.

14. J. Piaget, *The Origins of Intelligence in Children* (Norton, New York, 1952).

15. M. A. Arbib, E. J. Conklin and J. Hill, *From Schema Theory to Language* (Oxford University Press, Oxford, 1987).

16. J. L. McClelland and T. T. Rogers, The parallel distributed processing approach to semantic cognition, *Nature Rev. Neurosci.* **4** (2003) 310–322.

17. T. T. Rogers and J. L. McClelland, *Semantic Cognition: A Parallel Distributed Processing Approach* (MIT Press, Cambridge, MA, 2004).

18. W. J. Freeman, A neurobiological theory of meaning in perception, in *Proc. Int. Joint Conf. Neural Networks*, Portland, OR, 20–24 July 2003 (IEEE Press, Piscataway, NJ, 2003), pp. 1373–1378.

19. P. R. Cohen and C. R. Beal, Natural semantics for a mobile robot, Technical Report 00-59, University of Massachusetts Computer Science (2000).

20. T. Ziemke, Rethinking grounding, in *Understanding Representation in the Cognitive Sciences: Does Representation Need Reality?*, eds. A. Riegler, A. von Stein and M. Peschl (Kluwer Academic/Plenum Press, New York, 1999), pp. 177–199.

21. S. Harnad, The symbol grounding problem, *Physica D* **42** (1990) 335–346.

22. F. Rieke, D. Warland, R. de Ruter van Steveninck and W. Bialek, *Spikes: Exploring the Neural Code*, 1st edn. (MIT Press, Cambridge, MA, 1997).

23. Y. Choe and N. H. Smith, Motion-based autonomous grounding: Inferring external world properties from internal sensory states alone, in *Proc. 21st Nat. Conf. Artificial Intelligence*, Boston, MA, 16–20 July 2006, eds. Y. Gil and R. Mooney (AAAI Press, Menlo Park, CA, 2006), pp. 936–641.

24. D. Marr and E. Hildreth, Theory of edge detection, *Proc. Roy. Soc. London Ser. B* **207** (1980) 187–217.

25. J. G. Daugman, Two-dimensional spectral analysis of cortical receptive field profiles, *Vision Res.* **20** (1980) 847–856.

26. N. Petkov and P. Kruizinga, Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: Bar and grating cells, *Biol. Cybern.* **76** (1997) 83–96.

27. M. C. Doyle and R. Walker, Multisensory interactions in saccade target selection: Curved saccade trajectories, *Exper. Brain Res.* **142** (2002) 116–130.

28. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA, 1998).

29. R. S. Sutton, Learning to predict by the methods of temporal differences, *Mach. Learn.* **3** (1988) 9–44.

30. C. J. C. H. Watkins, Learning from delayed rewards, Ph.D. dissertation, University of Cambridge, England (1989).

31. N. H. Smith, Learning sensory invariance driven action in complex input environments (2007), under preparation.

32. R. L. Bishop and S. I. Goldberg, *Tensor Analysis on Manifolds* (Dover, New York, 1980) [original edition: Macmillan (1968)].

33. R. P. N. Rao and D. L. Ruderman, Learning lie groups for invariant visual perception, in *Advances in Neural Information Processing Systems 11*, eds. M. S. Kearns, S. A. Solla and D. A. Cohn (MIT Press, Cambridge, MA, 1999), pp. 810–816.

34. N. Almássy and O. Sporns, Perceptual invariance and categorization in an embodied model of the visual system, in *Biorobotics: Methods and Applications*, eds. B. Webb and T. R. Consi (AAAI Press/MIT Press, Menlo Park, CA, 2001), pp. 123–143.

35. J. J. Gibson, *The Perception of the Visual World* (Houghton Mifflin, Boston, 1950).

36. J. K. O'Regan and A. Noë, A sensorimotor account of vision and visual consciousness, *Behav. Brain Sci.* **24**(5) (2001) 883–917.

37. D. Philipona, J. K. O'Regan and J.-P. Nadal, Is there something out there? Inferring space from sensorimotor dependencies, *Neural Comput.* **15** (2003) 2029–2050.

38. D. Philipona, J. K. O'Regan, J.-P. Nadal and O. J.-M. D. Coenen, Perception of the structure of the physical world using unknown multimodal sensors and effectors, in *Advances in Neural Information Processing Systems 16*, eds. S. Thrun, L. Saul and B. Schölkopf (MIT Press, Cambridge, MA, 2004), pp. 945–952.

39. A. J. Bell, Levels and loops: The future of artificial intelligence and neuroscience, *Philos. Trans. Roy. Soc. London* **354** (1999) 2013–2020.

40. R. Platt Jr., A. H. Fagg and E. A. Grupen, Nullspace composition of control laws for grasping, in *Proc. 2002 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Lausanne, Switzerland, 30 September–4 October 2002 (IEEE Press, Piscataway, NJ, 2002), pp. 1717–1723.

41. G. Tevatia and S. Schaal, Inverse kinematics for humanoid robots, in *Proc. IEEE Int. Conf. Robotics and Automation 2000*, San Francisco, CA, 24–28 April 2000, Vol. 1, ed. O. Khatib (IEEE, Piscataway, NJ, 2000), pp. 294–299.

42. R. Held and A. Hein, Movement-produced stimulation in the development of visually guided behavior, *J. Comparat. Physiol. Psychol.* **56** (1963) 872–876.

43. P. Bach-y-Rita, *Brain Mechanisms in Sensory Substitution* (Academic Press, New York, 1972).
44. P. Bach-y-Rita, Tactile vision substitution: Past and future, *Int. J. Neurosci.* **19** (1983) 29–36.
45. R. A. Brooks, A robust layered control system for a mobile robot, *IEEE J. Robot. Autom.* **2**(10) (1986) 14–23.
46. R. A. Brooks, C. Braezeal (Ferrell), R. Irie, C. C. Kemp, M. Marjanović, B. Scassellati and M. M. Williamsom, Alternative essences of intelligence, in *Proc. 15th Natl. Conf. Artificial Intelligence*, Madison, WI, 26–30 July 1998, eds. J. Mostow and C. Rich (AAAI Press, Menlo Park, CA, 1998), pp. 961–976.
47. A. Edsinger and C. C. Kemp, What can I control? A framework for robot discovery, in *6th Int. Conf. Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, Paris, France (20–22 September 2006), online proceedings: http://www.csl.sony.fr/epirob2006/papers/EdsingerKemp.pdf.
48. K. Gold and B. Scassellati, Learning acceptable windows of contingency, *Connect. Sci.* **18** (2006) 217–228.
49. P. R. Cohen and C. R. Beal, Natural semantics for a mobile robot, University of Massachusettes, Department of Computer Science, Technical Report 00-59 (2000).
50. J. M. Mandler, How to build a baby: On the development of an accessible representational system, *Cogn. Dev.* **3** (1992) 113–136.
51. D. M. Pierce, Map learning with uninterpreted sensors and effectors, Ph.D. dissertation, Department of Computer Sciences, The University of Texas at Austin, Austin, TX (1995).
52. D. Pierce and B. Kuipers, Map learning with uninterpreted sensors and effectors, *Artifi. Intell.* **92** (1997) 169–227.
53. M. Bowling, A. Ghodsi and D. Wilkinson, Action respecting embedding, in *Proc. 22nd Int. Conf. Machine Learning*, Bonn, Germany, 7–11 August 2005, eds. L. De Raedt and S. Wrobel (Omni Press, Madison, WI, 2005), pp 65–72.
54. S. Singh, A. G. Barto and N. Chentanez, Intrinsically motivated reinforcement learning, in *Advances in Neural Information Processing Systems 17*, eds. L. K. Saul, Y. Weiss and L. Bottou (MIT Press, Cambridge, MA, 2005), pp. 1281–1288.
55. Y. Choe and R. Miikkulainen, Contour integration and segmentation in a self-organizing map of spiking neurons, *Biol. Cybern.* **90** (2004) 75–88.
56. Y. Choe and H.-F. Yang, Co-development of visual receptive fields and their motor primitive-based decoding scheme, in *Society for Neuroscience Abstracts*, Society for Neuroscience, Washington, DC (2006) Program No. 734.6. Online.
57. D. LaBerge, *Attentional Processing* (Harvard University Press, Cambridge, MA, 1995).
58. N. Misra, Comparison of motor-based versus visual representations in object recognition tasks, Master's thesis, Department of Computer Science, Texas A&M University, College Station, Texas (2005).
59. F. C. T. Moore, *Bergson: Thinking Backwards* (Cambridge University Press, Cambridge, UK, 1996).
60. S. Harnad, personal communication (2005).
61. R. Kozma and W. J. Freeman, Basic principles of the KIV model and its application to the navigation problem, *J. Int. Neurosci.* **2** (2003) 125–145.
62. E. Naito, P. E. Roland and H. H. Ehrsson, I felt my hand moving: A new role of the primary motor cortex in somatic perception of limb movement, *Neuron* **36** (2002) 979–988.
63. M. Bensafi, J. Porter, S. Pouliot, J. Mainland, B. Johnson, C. Zelano, N. Young, E. Bremner, D. Aframian, R. Khan and N. Sobel, Olfactomotor activity during imagery mimics that during perception, *Nat. Neurosci.* **6** (2003) 1142–1144.

64. A. Bompas, J. J. Clark and J. K. O'Regan, Colour perception in the sensorimotor contingency theory, *Perception* **31** (supplement) (2002) 90 [*Proc. 25th European Conf. Visual Perception, Glasgow, Scotland, 25–29 August 2002*].
65. E. T. Rolls, Representation in the brain, *Synthese* **129** (2001) 153–171.
66. J. M. Fuster, The prefrontal cortex — An update: Time is of the essence, *Neuron* **30** (2001) 319–333.
67. G. Rizzolatti, L. Fadiga, V. Gallese and L. Fogassi, Premotor cortex and the recognition of motor neurons, *Cogn. Brain Res.* **3** (1996) 131–141.
68. G. Rizzolatti, L. Fogassi and V. Gallese, Neurophysiological mechanisms underlying the understanding and imitation of action, *Nat. Rev. Neurosci.* **2** (2001) 661–670.
69. Y. Choe, A deeper semantic role for the mirror system, *Behavioral and Brain Sciences*, Vol. 28 (2005), (On-line supplemental commentary on Arbib (2005) *Behavioral and Brain Sciences*, Vol. 28, pp. 105–167. URL: http://www.bbsonline.org/Preprints/Arbib-05012002/Supplemental.) 1 page.
70. P. Bach-y-Rita, Tactile sensory substitution studies, *Ann. New York Acad. Sci.* **1013** (2004) 83–91.
71. G. Kampis, The natural history of agents, in *Agents Everywhere*, eds. L. Gulyás, G. Tatai and J. Váncza (Springer, Budapest, 1999), pp. 24–48.
72. G. G. Turrigiano, Homeostatic plasticity in neuronal networks: The more things change, the more they stay the same, *Trends Neurosci.* **22**(5) (1999) 221–227.
73. G. G. Turrigiano, K. R. Leslie, N. S. Desai, L. C. Rutherford and S. B. Nelson, Activity-dependent scaling of quantal amplitude in neocortical neurons, *Nature* **391** (February 1998) 845–846.
74. T. J. Sullivan and V. R. de Sa, Homeostatic synaptic scaling in self-organizing maps, *Neural Networks* **19** (2006) 734–743.
75. Y. Choe and S. K. Bhamidipati, Autonomous acquisition of the meaning of sensory states through sensory-invariance driven action, in *Biologically Inspired Approaches to Advanced Information Technology: First International Workshop*, eds. A. J. Ijspcert, M. Murata and N. Wakamiya, Lecture Notes in Computer Science, Vol. 3141 (Springer, Berlin, 2004), pp. 176–188.

Yoonsuck Choe is an Assistant Professor in the Department of Computer Science at Texas A&M University. He received his B.S. degree in Computer Science from Yonsei University (Korea) in 1993, and his M.S. and Ph.D. degrees in Computer Science from the University of Texas at Austin in 1995 and 2001. His current research areas include computational neuroscience, computational neuroanatomy, computational and biological vision, and autonomous semantics in sensorimotor agents.

Huei-Fang Yang is a Ph.D. student in the Department of Computer Science at Texas A&M University. She received her B.Ed. and M.Ed. degrees in Information and Computer Education from National Taiwan Normal University (Taiwan) in 1998 and 2001. Her research interests include computer vision and image processing.

Daniel Chern-Yeow Eng is currently pursuing his M.Sc. degree in Computer Science at Texas A&M University. He received his B.S degree in Computer Engineering from Texas A&M in 2006. His current research interests include machine learning, computational and biological vision, and artificial intelligence.