frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 19 April 2016
doi: 10.3389/frobt.2016.00022

A Self-Organized Internal Models Architecture for Coding Sensory–Motor Schemes

Esau Escobar-Juárez[1], *Guido Schillaci*[2], *Jorge Hermosillo-Valadez*[1]* and *Bruno Lara-Guzmán*[1]

[1] *Centro de Investigación en Ciencias-(IICBA), Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, México,* [2] *Adaptive Systems Group, Department of Computer Science, Humboldt-Universität zu Berlin, Berlin, Germany*

Cognitive robotics research draws inspiration from theories and models on cognition, as conceived by neuroscience or cognitive psychology, to investigate biologically plausible computational models in artificial agents. In this field, the theoretical framework of Grounded Cognition provides epistemological and methodological grounds for the computational modeling of cognition. It has been stressed in the literature that *simulation*, *prediction*, and *multi-modal integration* are key aspects of cognition and that computational architectures capable of putting them into play in a biologically plausible way are a necessity. Research in this direction has brought extensive empirical evidence, suggesting that *Internal Models* are suitable mechanisms for sensory–motor integration. However, current Internal Models architectures show several drawbacks, mainly due to the lack of a unified substrate allowing for a true sensory–motor integration space, enabling flexible and scalable ways to model cognition under the embodiment hypothesis constraints. We propose the Self-Organized Internal Models Architecture (SOIMA), a computational cognitive architecture coded by means of a network of self-organized maps, implementing coupled internal models that allow modeling multi-modal sensory–motor schemes. Our approach addresses integrally the issues of current implementations of Internal Models. We discuss the design and features of the architecture, and provide empirical results on a humanoid robot that demonstrate the benefits and potentialities of the SOIMA concept for studying cognition in artificial agents.

Keywords: internal models, cognitive robotics, self-organized maps, sensory–motor schemes, computational architecture

OPEN ACCESS

Edited by:
Francesco Becchi,
Telerobot Labs s.r.l., Italy

Reviewed by:
Arnaud Blanchard,
University of Cergy-Pontoise, France
Felix Reinhart,
Bielefeld University, Germany

****Correspondence:***
Jorge Hermosillo Valadez
jhermosillo@uaem.mx

Specialty section:
This article was submitted to Humanoid Robotics, a section of the journal Frontiers in Robotics and AI

Received: *02 October 2015*
Accepted: *04 April 2016*
Published: *19 April 2016*

Citation:
Escobar-Juárez E, Schillaci G, Hermosillo-Valadez J and Lara-Guzmán B (2016) A Self-Organized Internal Models Architecture for Coding Sensory–Motor Schemes. Front. Robot. AI 3:22. doi: 10.3389/frobt.2016.00022

1. INTRODUCTION

Cognitive robotics is an active research field in the cognitive sciences since the role of embodiment has been acknowledged as crucial to understand and reproduce natural cognition, showing as well a stance against the classic theory of cognition as symbolic processing. Research in cognitive robotics draws inspiration from theories and models on cognition, as conceived by neuroscience or cognitive psychology, to investigate biologically plausible computational models in artificial agents. Scientific aims include studying the implications of these models under controlled conditions and providing agents with basic cognitive skills (Pfeifer and Scheier, 2001).

In this research field, Grounded Cognition (Barsalou, 2008) constitutes a theoretical reference framework, including the account of embodied cognition, which stresses the importance of the body–environment interaction for the structuring and emergence of cognitive skills (Wilson, 2002).

Under this perspective, we are committed to investigate biologically plausible computational architectures in which to model cognition effectively. This issue has not trivial answers since there are several constraints on the nature, the role, and the architectural integration of the underlying artificial mechanisms by means of which we shall achieve computationally effective ways to model cognition. Some of these most relevant constraints are revised now.

In Grounded Cognition, all aspects of experience, perceptual states (for instance, those produced by vision, hearing, touching, tasting), together with internal bodily states and action, have neural correlates in the brain that are stored in memory. These neural activation patterns constitute multi-modal representations that are re-enacted during perception, memory, and reasoning. Modal re-enactments of these patterns constitute *internal simulation* processes (Barsalou, 2003) and are considered to lie at the heart of the off-line characteristics of cognition (Wilson, 2002). Thus, the theory of simulation is at the core of the embodied cognition hypothesis.

This shift in the paradigm about cognition has necessarily brought new design considerations on computer models in order to achieve embodied or grounded cognition, which have taken center stage in Artificial Intelligence [e.g., Grush (2004), Svensson and Ziemke (2004), and Pezzulo et al. (2011, 2013a)]. Thus, emphasis has been made on the predictive learning and internal modeling capabilities of the sensory–motor system (Pezzulo et al., 2013b).

Furthermore, in the embodied cognition framework, the acquisition of sensory–motor schemes is central (Pfeifer and Bongard, 2007), for they underlie cognition (Lungarella et al., 2003) and are grounded in the regularities of the sensory–motor system interactions with its environment. The cerebral cortex provides the necessary substrate for the development of these sensory–motor schemes as it constitutes the locus of the integration of multi-modal information.

All these considerations point to the fact that *simulation*, *prediction*, and *multi-modal integration* are key aspects of cognition and it has been stressed in the literature the necessity to achieve cognitive architectures capable of putting them into play in a biologically plausible way (Pezzulo et al., 2011). This paper is an attempt in this direction.

Based on extensive empirical evidence of its putative functionality in the Central Nervous System (Kawato, 1999; Blakemore et al., 2000; Wolpert et al., 2001), *Forward* and *Inverse Models* provide arguably a sound epistemological basis to understand cognitive processes at a certain level of description under the embodied cognition framework.

A thorough review of the implementations of internal models is out of the scope of this work. However, it is worth noting that most of the implementations show shortcomings in light of our previous discussion [e.g., see Arceo et al. (2013)]. First, we find that current implementations of internal models lack of flexibility as a consequence of the computational tools used. This translates into the fact that learning plasticity is highly reduced or even absent [e.g., see Lara and Rendon-Mancha (2006), Dearden (2008), Möller and Schenck (2008), and Schenck et al. (2011)]. Second, the implementations redound in *ad hoc* inverse and forward models, not easily scalable, and in some cases, using different networks for different motor commands [e.g., Möller and Schenck (2008)]. Finally, in the literature, there is an abstract and high-level coding of inverse models as in Dearden (2008), where inverse models are coded as direct actions.

We propose a new computational architecture for building cognitive tasks under the paradigm of Grounded Cognition: the Self-Organized Internal Models Architecture (SOIMA), a computational cognitive architecture coded by means of a network of self-organized maps, implementing coupled internal models that allow multi-modal associations. The SOIMA tackles integrally the issues of current implementations as will be discussed in the sequel.

The structure of the paper is as follows: in Section 2, we introduce the SOIMA architecture, explaining its theoretical foundations, justifying the Internal Models approach for modeling cognition and detailing the SOIMA's structure. Also, the features that make of it a suitable cognitive architecture tackling current implementations' shortcomings are introduced in this section. In Section 3, we provide two experimental case studies to demonstrate the architecture's functionality. We first introduce a case study for saccadic control in order to demonstrate the SOIMA features in detail. Then, we show a Hand–Eye Coordination task allowing us to demonstrate a scaling-up of the architecture, showing how the connectivity enhancements enable flexible and effective ways to model more complex tasks. In Section 4, we conclude by discussing the results and perspectives for future research.

2. SELF-ORGANIZED INTERNAL MODELS ARCHITECTURE: SOIMA

2.1. Biological Foundations

Brain plasticity regulates our capability to learn and to modify our behavior. Plastic changes are induced in neural pathways and synapses by the bodily experience with the external environment.

In the neurosciences literature, it has been proposed that the rich multi-modal information flowing through the sensory and motor streams is integrated in a sort of body schema, or body representation. Fundamental for action planning and for efficiently interacting with the environment (Hoffmann et al., 2010), such a body representation would be acquired and refined over time, already during pre-natal developmental stages.

For example, Rochat (1998) showed that infants exhibit, at the age of 3 months, systematic visual and proprioceptive self-exploration. The authors also report that infants, by the age of 12 months, possess a sense of a calibrated intermodal space of their body, that is a perceptually organized entity which they can monitor and control (Rochat and Morgan, 1998). As discussed by Maravita et al. (2003) and Maravita and Iriki (2004), converging evidence from animal and human studies suggests that the primate brain constructs various body-part-centered representations of space, based on the integration of different motor and sensory signals, such as visual, tactile, and proprioceptive information.

Sensory receptors and effector systems seem to be organized into topographic maps that are precisely aligned both within and across modalities (Udin and Fawcett, 1988; Cang and Feldheim, 2013). Such topographic maps self-organize throughout the brain development in a way that adjacent regions process spatially close sensory parts of the body. Kaas (1997) reports a number of studies showing the existence of such maps in the visual, auditory, olfactory, and somatosensory systems, as well as in parts of the motor brain areas.

All this evidence suggests thus that cognition relies on self-organized body-mapping structures integrating sensory–motor information. But how does this integration takes place?

The work of Damasio (1989) and Meyer and Damasio (2009) proposes a functional framework for multi-modal integration supported by the theory of convergence-divergence zones (CDZ) of the cerebral cortex. This theory holds that specific cortical areas can act as sets of pointers to other areas and, therefore, relate various cortical networks to each other.

CDZ integrate low level cortical area networks (close to the sensory or motor modalities) with high level amodal constructs, which solves the problem of multi-modal integration since it enables the extraction of complex pure and non-segmented sensory information units and sensory–motor contingencies.

In the CDZ convergence process, modal information spreads to the multi-modal integration areas; while in the divergence process, multi-modal information propagates to modal networks generating the re-enactment of sensory or motor states. It is in this sense that the propagation bi-directionality provides the mechanism of mental imagery and the re-enactment of sensory–motor states.

This bi-directional capability is, thus, fundamental for multi-modal integration and, hence, for cognition. The lack of this property is precisely one of the main limitations of current cognitive architectures that our model tackles as will be shown in subsequent sections.

2.2. Internal Models

Internal models merge in a natural way sensory and motor information and create a multi-modal representation (Wolpert and Kawato, 1998). These models also provide agents with anticipation, prediction, and motor planning capabilities by means of internal simulations (Schillaci et al., 2012b).

We are particularly interested in the pair formed by Inverse-Forward models. The inverse model (IM) is a controller (**Figure 1A**), which generates the motor command (M_t^*) needed to achieve a desired sensory state (S_{t+1}) given a current sensory state (S_t). The forward model (FM) is a predictor (**Figure 1B**) that predicts the sensory state entailed (S_{t+1}^*) by some action of the agent (M_t) given a current sensory state (S_t).

While the IM is mainly required for motor control, the FM has been proposed as a possible model for a number of important issues, among which are sensory cancelation (Blakemore et al., 2000), state estimation (Wolpert et al., 1995), and body map acquisition (Schillaci et al., 2012a).

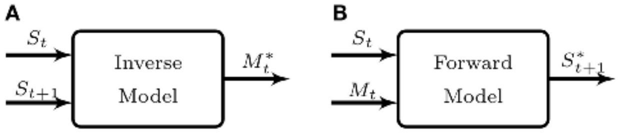

FIGURE 1 | Internal models canonical representation. **(A)** Inverse model: given the current sensory state St and a desired sensory state for time t + 1, St+1 this model suggests the motor command M_t^* that brings about the change in the sensory state of the agent. **(B)** Forward model: given the current sensory state S_t and a motor command M_t, the forward model predicts the resulting *sensory state prediction* S_{t+1}^*, were the command executed. It is worth noting the difference between S_{t+1}^* and S_{t+1}. The former is the actual expected state for the next time-step, but the latter may require several time-steps in order to be attained.

The coupled pair IM–FM has been introduced by Jordan and Rumelhart (1992) from control theory. In neuroscience, one of the first proposals was the MOSAIC architecture by Wolpert and Kawato (1998) and has been used in action recognition (Demiris and Khadhouri, 2006; Arceo et al., 2013), own body distinction (Schillaci et al., 2013), and mental simulation (Möller and Schenck, 2008).

In Cognitive Robotics, internal models have been used for action execution and recognition (Dearden, 2008), safe navigation planning (Lara and Rendon-Mancha, 2006; Möller and Schenck, 2008), and saccades control (Schenck et al., 2011). On the other hand, several works have proposed IM–FM couplings to perform different tasks. For example, in Schillaci et al. (2012b), several IM–FM pairs are used to recognize an action when comparing the output of each pair with the real situation. In the case of Schenck (2008), each pair is used in order to produce the motor command enabling an agent to reach a desired position, where the FM acts as the desired position monitor.

Internal Models are thus a suitable mechanism for multi-modal representations. They constitute a sound basis for modeling cognition and they also provide a coherent epistemological framework for studying it under the embodied cognition framework.

In our work, we propose an architecture that preserves the structural ideas put forward by Damasio along with the self-organizing and multi-modal integration properties of the brain, in the framework of internal models. This allows for building a mechanism for the integration and generation of multi-modal sensory–motor schemes in the framework of Grounded Cognition.

The SOIMA relies on two main learning mechanisms. The first one consists in Self-Organizing Maps (SOMs) that create clusters of modal information coming from the environment. The second one codes the internal models by means of connections between the first maps using Hebbian learning. This Hebbian association process generates sensory–motor patterns that represent actual sensory–motor schemes.

This coding approach of internal models using SOMs and Hebbian learning allows for a modular implementation, and constitutes the main contribution of our work. The architecture allows for an integrated learning strategy and provides means for building coupled sensory–motor schemes in a flexible way. Most of the previous approaches of internal models implementations, as reported in the literature, provide different computational substrates that connect to each other in order to synthesize the coupled model. The substrates may even be of different nature (e.g., different kinds of neural networks), which obligates to use distinct learning strategies for each model. In many cases also, the resulting models are *ad hoc* for the task. Our approach synthesizes the coupled model on the same substrate, conferring connectivity enhancements that allow for modular and flexible internal models implementations and sensory–motor scheme maps. This mapping capability may be exploited in interesting ways as will be discussed in the conclusions.

We now discuss the implementation details of these two learning mechanisms in the SOIMA.

2.3. SOIMA's Structure

2.3.1. Modal Information Clustering

In the SOIMA, the basic units are SOMs (Kohonen, 1990) that generate clusters of information coming from different modalities (sensory or motor) or from other SOMs.

When training a SOM, a topological organization occurs in a space of lower dimension (2D or 3D) than the modal input space. This organization corresponds to a partition of the input space into regions that reveal the similarities of the input data.

A SOM is an artificial neural network endowed with an unsupervised learning mechanism based on vector quantization. Vector quantization refers to the fitness of a probability density function to a discrete set of prototype vectors.

In our case, we are interested in evaluating the differences between the vectors not only in terms of their relative distance but also in terms of their orientation. The cosine similarity can be used to obtain the differences in orientation, so we use both measures for clustering data in the SOM. Thus in our design, when a vector x occurs at the input, the activation A_j of each node in the SOM is defined as

$$A_j = \frac{1}{2}\left(\left\|\mathrm{x} - \mathrm{w}_j\right\| + 1 - \frac{\mathrm{x}\cdot\mathrm{w}_j}{\left\|\mathrm{x}\right\|\left\|\mathrm{w}_j\right\|}\right) \tag{1}$$

where w_j is the vector of weights between the input and the node j, the first term is the Euclidean distance between x and w_j and the second term is the cosine similarity. The winning node is the one with the lowest activation.

Once the winning node is computed, the weights of the neurons are updated according to the following equation

$$\Delta\mathrm{w}_j = \alpha(t)h_j(\mathrm{x} - \mathrm{w}_j) \tag{2}$$

where w_j is the weight between the input vector x and node j, h_j is the neighborhood function of node j defined as

$$h_j = e^{(\frac{-\beta_j}{2\sqrt{n}})} \tag{3}$$

where β_j is the distance between node j and the winning unit, and n is the total number of nodes in the map. If β_j is greater than the size of the neighborhood ν then $h_j = 0$, ν decreases monotonically from ν_i[1] to $\nu_f = 1$, where ν_i and ν_f are the initial and final neighborhood sizes, respectively.

And finally $\alpha(t)$ is the learning rate that increases as a function of time, defined by:

$$\alpha(t) = \alpha^i + \left(\frac{t(\alpha^f - \alpha^i)}{(\nu_i - \nu_f)}\right) \tag{4}$$

where α^i and α^f are the initial and final learning rates, respectively, and t is the current learning period.

2.3.2. Modal Maps Association

The association between SOMs of different modalities has been reported in recent work. In Westerman and Miranda (2002), the association between vocalization and hearing maps can be used for modeling the emergence of vocal categories. In Li et al. (2007), clues about the vocabulary development age in infants, using a similar association scheme, were found. In Mayor and Plunkett (2010), an association between visual and hearing maps was used to determine the taxonomic response in early learning of words. Morse et al. (2010a) integrates different sensory and motor modal maps through a changing network with Hebbian learning to build a semantic meaning acquisition system.

In our work, we create the modal association between different SOMs through weights connected using the well-known Hebbian learning rule (Hebb, 1949). The rule states

> When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased (Hebb, 1949).

In this respect, the Hebbian rule establishes that the connection between neurons is reinforced according to the activation of neurons that participate in the connection. In our model, we use the following positive Hebbian rule for modulating connections between nodes of different maps:

$$\Delta w_{ij} = \alpha u_i u_j \tag{5}$$

where w_{ij} is the weight of the connection between the node the node i, and j α is the learning rate, u_i is the activation of the node i as u_j is the activation of the node j.

Figure 2 depicts the proposed architecture showing two maps (S and M) corresponding to the modalities of the agent. We consider M as a modality so that, together with S, the top map forms a multi-modal representation (MMR). The idea of Hebbian training is to modulate a network of connections between the top SOM and each modality SOM. Each node of the top SOM is connected to every node of the modality SOMs. These connections are originally set to 0.

The association process takes place as follows. The two maps S and M are fed with the sensory and motor data generated

[1] This initial value is oftentimes taken as the 15% of the size of the map.

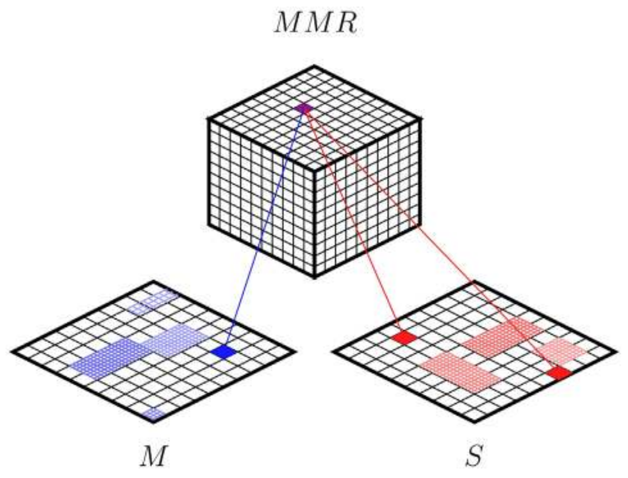

FIGURE 2 | **Self-organized internal models architecture (SOIMA).** Two modality maps are shown representing *M* and *S*. The map at the top integrates the modalities coding, thus, sensory–motor schemes in a Multi-Modal Representation (*MMR*).

throughout the interaction of the agent with the environment. Every time an input pattern is introduced, there is a winning node in *S* and *M*, respectively. Then the inputs to the top map are the coordinates of these winning units, so that a corresponding winning unit occurs at the top SOM. A Hebbian modulation is then applied to the connection between these nodes. In this way, a sensory–motor scheme is coded on the *MMR* through the Hebbian mechanism. Once the system has being trained, this association allows for retrieving all the modalities when any of them is present.

As can be seen in **Figure 2**, each winning unit of the *MMR* map receives one connection coming from the *M* map and two connections coming from the *S* map, representing two different time steps (a change in the sensory situation); the motor command is the one associated with that change in the sensory situation. Thus, the trained system associates a triplet formed by a sensory situation, a motor command, and a corresponding predicted sensory situation associated with these two. It could be said then that the *MMR* codes the associated triplet, as each node in this map codes for a specific sensory–motor experience of the agent. The *MMR* map is coded as a cube in order to better represent the multi-modal space. In this configuration, each triplet has 26 direct neighbors providing a richer structure.

We now discuss the main attributes of the SOIMA.

2.4. SOIMA's Features

One of the main advantages of the SOIMA resides in its bi-directional functionality, since it can work either as a forward or as an inverse model, depending on the inputs that are fed to the system. In other words, the system integrates the sensory–motor scheme in such a way that it is now independent of the directionality of the modal information flow.

When a forward model is required, an *S* and *M* signal for the time *t* should be present, activating the corresponding maps and their connections toward the *MMR* of the system[2] (see **Figure 2**). Thence, the signal would spread back to the map *S*, producing with its activation the prediction of the sensory state at the time $t + 1$.

If, on the contrary, an inverse model is needed, then the required inputs are two sensory situations, coded in *S* corresponding to times *t* and $t + 1$, in turn triggering the activation of the *MMR* map and, thence, activating the node in *M* corresponding to the pair in *S*.

Some interesting features of the *MMR* are noteworthy. The *MMR* allows the bi-directional feature to be functional all the time. In other words, any model (IM or FM) can be easily implemented by instantiating the corresponding inputs for the required functionality. In this way, both internal models are coded on the same substrate, enabling the design of integrated learning strategies.

Moreover, the *MMR* allows for the construction of coupled IM–FM pairs in a modular way. Indeed, as either model (IM or FM) can be instantiated at any time, the output of one of them can be used as input to the other, constituting thus an IM–FM coupling. Thus, several IM–FM couplings can be instantiated sequentially, so as to build a simulation process for instance. Hence, it is possible to feed sequentially in time the IM or FM, either with data coming from the environment or produced by the system itself. Also, the *MMR* allows for the integration of other *MMR*s, which enables the coupling of distinct sensory–motor schemes. These features will be illustrated in the experiments.

Last, but not least, the *MMR* is not built from an abstract representation, in the sense of being defined by the programmer of the system as in the classical AI paradigm. Rather, it constitutes a representation *grounded* in the bodily constrained interaction of the agent with its surrounding environment.

The online learning ability of the architecture is also noteworthy. While running, the system is able to acquire new knowledge in order to improve its performance on an ongoing basis. This capability is achieved through the updating of the weights interconnecting the SOMs, so the system can adapt to unfamiliar situations as they arrive. This feature will be made clearer in the next section.

3. EXPERIMENTS AND RESULTS

In order to demonstrate the functionality of the SOIMA, we introduce two case studies using a NAO[3] humanoid robot. The first experiment is intended to demonstrate the SOIMA's functionality by modeling saccadic movements of the eye in order to center a stimulus. This experiment was designed to describe the detailed workings of the SOIMA implementation and shows also that the SOIMA approach allows to cope with typical dimensionality issues of the visual input space. The second case

[2] The activation of the top map constitutes the integration and multi-modal representation of the event.

[3] Developed by Aldebaran Robotics.

study introduces an example of how the SOIMA can be scaled upwardly to model more complex cognitive tasks. In this second experiment, we aim at implementing a Hand–Eye Coordination (*HEC*) strategy using the SOIMA. Here, the architecture is used as a building block, allowing for exploiting previously acquired knowledge.

3.1. Saccadic Control

3.1.1. Visuomotor Schemes Modeling

Rapid eye motions, the so-called *saccadic movements* (Leigh and Zee, 1999), are intended to project the image of the visual area of interest in the most sensitive part of the retina, called the fovea. Saccadic control is a canonical problem involving sensory predictive and fine-tuning motor control capabilities.

It is worth mentioning a brief comment on the work of Kaiser et al. (2013) that introduces a saccadic control system based on internal models and asserts that addressing the image prediction problem is rather highly complex due to the input high dimensionality. For this reason, in their proposal predictions are not made, rather inverse mappings are computed from the output image to the input image.

The implementation of internal models can be made with any of the available learning techniques in artificial intelligence. However, there are serious limitations to the use of images, because they increase the dimensionality of the problem as well as the difficulty on finding regularities in the inputs. As an end result, the coding and learning of visuomotor schemes becomes a major difficulty.

By contrast, the scheme that we propose addresses the prediction problem using images of high dimensionality, which enables the development of more versatile models. Our system allows for learning the relationship between the camera motions and the corresponding sensory changes. Once learned, the model works as a mechanism that retrieves the motor command required to take some stimulus in the image to any desired area in the same image. In particular, we want the model to focus on some salient stimulus. A similar implementation is presented in Karaoguz et al. (2009) where a SOM is used for gaze fixation.

We use as input an image grabbed from one of the cameras in the NAO humanoid robot. Our experiment on saccadic control is an instance of a modular system that implements different coupled internal models pairs (FM–IM). We use the simulations provided by the FM–IM pair to provide the motor commands necessary to place some stimulus present in the image at any specified position. In **Figure 3,** we show the schematic functional diagram of the system.

At this point, it is worth mentioning that an important feature of the architecture is the ability to learn the Hebbian connections online. Initially, there is no association between the maps in the system, i.e., all connections have a value of 0; therefore, no motor command may be suggested aiming at focusing the S_t stimulus by means of the inverse model. Hence, when a motor command cannot be retrieved from the system, a motor-babbling mechanism generates a random movement. This command is then executed, obtaining thus the S_{t+1} image. This information is integrated into the connections between SOMs using on-line Hebbian learning.

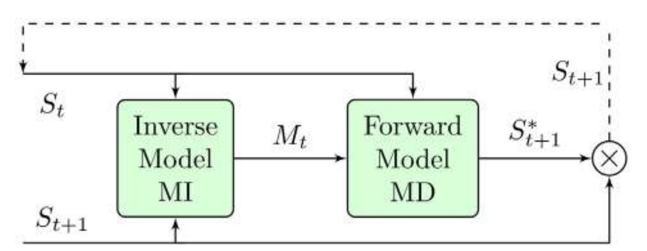

FIGURE 3 | **Saccadic control integrated model (explanation in text).**

The final result is the association of two nodes from the S map, one representing S_t and one representing S_{t+1} with one node in the M map representing the motor command that brings about this change in the sensory space.

Thus, in the first part of the forward-inverse coupling, input S_t represents the current visual sensory input, i.e., the image with a stimulus appearing at some arbitrary position. Input S_{t+1} represents the desired visual sensory state, i.e., the image with the stimulus in the desired position (this image is built or taken from a database). With these inputs, the inverse model suggests an initial motor command M_t^* aiming at the desired sensory change[4].

This motor command along with the image S_t is fed to the forward model to predict the sensory outcome S_{t+1}^*. This predicted image is compared with the desired one S_{t+1} to compute the error. In turn, the error is used to decide whether this output should be fed back as S_t, so that a corrective saccadic movement can be performed. In other words, supplementary control commands may be stacked together with the first M_t^* in order to reach the desired situation S_{t+1}.

Finally, once the motor commands required to bring S_t to the desired S_{t+1} are found, they are executed in the system with actual movements.

The details of the clustering of the modal maps and their Hebbian connections are discussed now.

Experiments were carried out on a simulated NAO humanoid robot endowed with 21 degrees of freedom (DOF) and two cameras. The Webots v8.0.3 was used to test the saccadic movements system. The experimental setup consisted in the robot facing a wall, situated at a distance of 40 cm, where a single visual stimulus was displayed, as shown in **Figure 4**. The model showed in **Figure 3** was used to control the saccadic movements toward the visual stimulus.

The two DOF associated to the agent's head movement were used: *yaw* (rotation around the vertical axis) and *pitch* (rotation around the horizontal axis); the camera is situated in the upper part of the head and has an image resolution of 640 × 480 pixels.

3.1.2. Sensory Input Processing

Learning requires a set of training patterns, each containing an image for S_t, one for S_{t+1} and a motor command that brings the

[4] Recall that once the system is trained, it can be used as an inverse or a forward model.

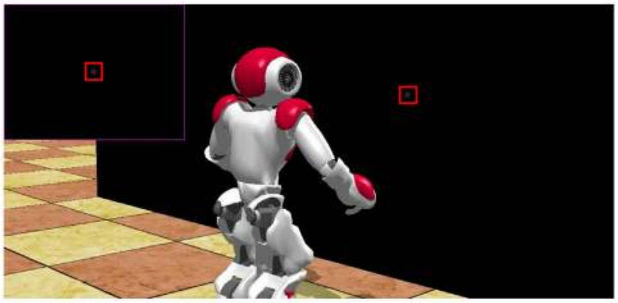

FIGURE 4 | **Experimental setup**.

system from S_t to S_{t+1}. The motor command is defined as a pair (Δ Yaw, Δ Pitch).

To build the images, two stages of processing are necessary. In the first stage, a fovealized image is obtained from the original camera image (640 × 480 pixels). To foveate the images, we apply a radial mapping with high resolution toward the center of the image and low toward the periphery. The mapping emulates the human retina properties containing high concentration of photoreceptors in the fovea. It fulfills a special function in our design, because when a stimulus is located near the central part of the image (fovea) a small change in Yaw or Pitch corresponds to a large position change of the stimulus in the image. This enables a more accurate detection of the position of the stimulus nearby the central region of the image. As a result, the task of centering a stimulus in the image is more accurate. The fovealization algorithm delivers a 320 × 240 image.

The second processing phase is intended to facilitate the identification of *salient stimuli* in the image captured by the camera. This is achieved through binary thresholding and gaussian smoothing. At this stage, the image size is 40 × 30 pixels, reducing the sensory input space dimensionality. In **Figure 5,** we can see the visual stimuli at the different stages of processing.

In our system, a motor command M_t is a change (Δ) in the orientation (in degrees) on the horizontal and vertical axes of the robot's head. Any change between two positions depends on the resolution of the motors. This resolution in our case was built using a mapping mechanism similar to that applied to the visual modality. This motor mapping consisted in a variable yaw–pitch movement resolution, being this higher in the center than in the periphery.

To visualize the motor space, a motor resolution image (I_{RM}) was made from a total of 5000 head joints configurations. Initially, I_{RM} is an image where all its pixels are set to 0. Then for each position, the center of mass of the visual stimulus in the image is computed. According to the location of the center of mass, the value of the corresponding pixel in I_{RM} is increased. This gave rise to an intensity image where each pixel value is proportional to the number of positions in each location.

The I_{RM} image is depicted in **Figure 6**. As it can be seen, the highest visual stimulus density on camera positions is located in the central region of the image.

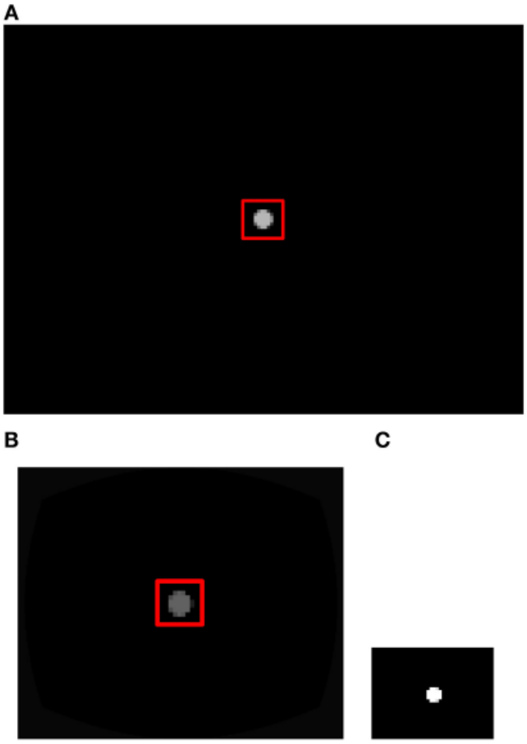

FIGURE 5 | Input image processing, (A) original image 640 × 480, (B) foveated image 320 × 240, and (C) saliency image 40 × 30.

3.1.3 SOMs Training

For the training of the SOMs, 5000 random patterns with different initial (S_t) and final (S_{t+1}) camera positions and their corresponding M_t were taken with the following structure:

- S_t: it was formed by a 1200 values vector normalized from 0 to 1, taken from a 40 × 30 pixel salient features image.
- S_{t+1}: it was formed with the same procedure as for S_t but with a different camera position.
- M_t: it consists of two values (ΔYaw, ΔPitch) built in accordance with the following:

$$\Delta Yaw = (Yaw_{S_t} - Yaw_{S_{t+1}}) \tag{6}$$

$$\Delta Pitch = (Pitch_{S_t} - Pitch_{S_{t+1}}). \tag{7}$$

normalized from 0 to 1. With the robot placed at 40 cm from the stimulus, the Yaw axis of the camera covers an angle of 43.5° and the Pitch axis 37.8°, assuring that the stimulus is always on sight. The value of 1 represents the biggest positive possible change above the corresponding axis and 0 represents the biggest possible change in the other direction. Values of (0.5, 0.5) represent absence of movement in both axes.

The SOIMA used is shown in **Figure 2**. We used a 30 × 30 SOM to code *S*, and a 40 × 40 for *M*, finally for the *MMR* a three-dimensional 30 × 30 × 30 SOM was used. These maps were trained individually using the respective collected modal information patterns using the procedure described in Section 2.3.

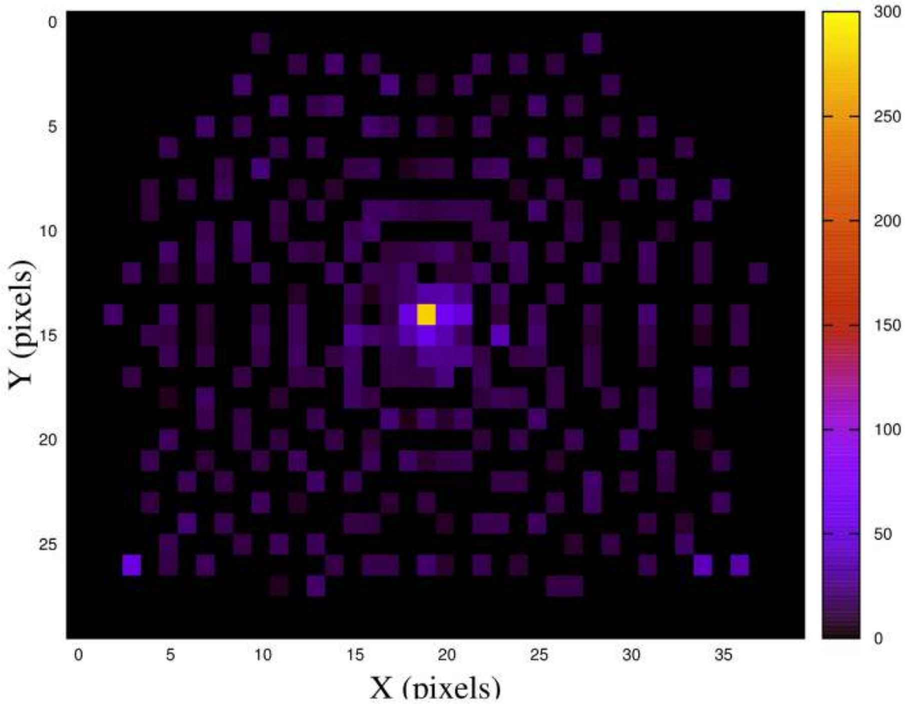

FIGURE 6 | **Visual stimulus distribution on different positions of the camera**.

3.1.4. Online Hebbian Learning

As mentioned in Section 3.1.1, the online learning capability gradually increases sensory–motor knowledge on saccadic movements, which decreases the use of the random mechanism (**Figure 7A**). In turn, the system reduces the error as it focuses the visual stimulus (**Figure 7B**).

Figure 7B depicts the quartiles of 11 subsets of the available data on the online learning error. Red lines show the medians of each subset. The figure shows that the variability of the error reduces with time. The mean value of each subset stabilizes quickly and falls within the third quartile, showing that the distributions of these data sets are not gaussian.

When the motor command is suggested by the system (i.e., when the SOIMA already contains a multi-modal association and is able to act as IM and FM), learning also occurs:

- when the architecture generates a sequence of two or more simulated saccadic movements to reach from S_t to S_{t+1}, a single motor command is learned as the association between the two sensory situations.
- in those cases, when after running a motor command the position error is greater than a certain threshold, another motor command is calculated and executed by re-enacting the system with S^*_{t+1} [5]. This new motor command is added to the previous motor command, so that this new association is integrated into the system as if it were a single execution.

3.1.5. Saccadic Control Execution

3.1.5.1. Prediction

For illustration purposes, we want the system to center the stimulus on the image in a foveation-like process.

In principle, training data covers the whole of the visual field of the camera, so it would be possible to give another location of the stimulus as desired sensory situation S_{t+1}. However, the desired stimulus S_{t+1} is built from a repository of images containing a single object in the center of the image.

A typical test example is depicted in **Figure 8**, in this two prediction steps are conducted:

- Step 1: the current camera image is fed into the system and processed according to areas of salient features to form S_t. The desired image S_{t+1} is also fed into, with the stimulus located at the center. These inputs go through the inverse model and generate a suggestion for the motor command M^*_t. The latter is in turn applied together with S_t to the forward model to generate a prediction S^*_{t+1}. This output is compared with the S_{t+1}, the desired situation, to compute the error.
- Step 2: if the error is greater than a threshold, S^*_{t+1} is fed back into the system again, as the new S_t, to obtain an additional motor command in order to achieve the marked position accurately.

It is worth noting that for both steps, the error is calculated between the prediction of the forward model and the desired

[5] The reader must bear in mind that the system codes the FM–IM coupling. In other words, it is possible to generate new motor commands from the FM predicted sensory output (S^*_{t+1}), by using it as input again to the IM. This new motor command can in turn be used once again to generate a new prediction.

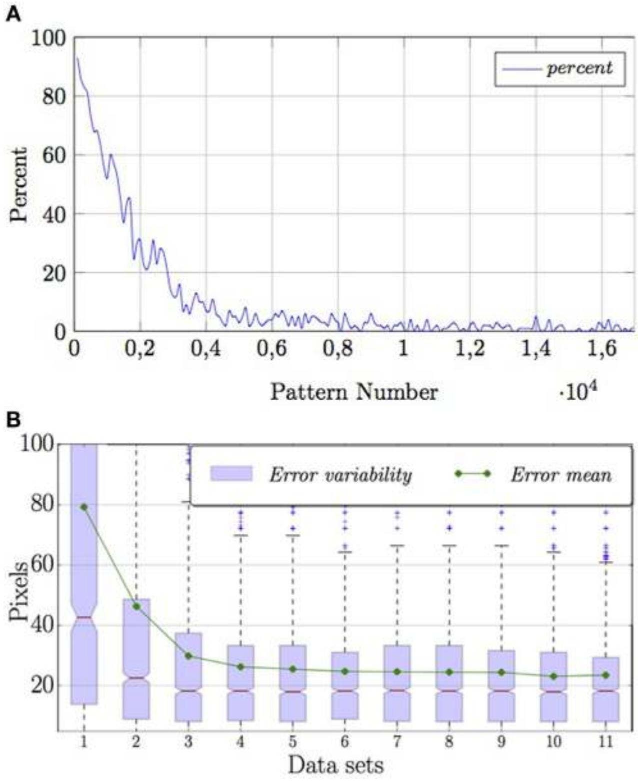

FIGURE 7 | **Performance of the system during training**. **(A)** Average percentage of random motor commands used as learning occurs. **(B)** Error in pixels between final position and center of image.

situation. Only when the error between the prediction and the desired state is lower than the threshold, a motor command is executed. This could mean that more than two internal simulations are run.

Ninety-five tests were performed on the saccadic control model (**Figure 9**) from different robot's head positions, with the initial stimulus located at distinct locations on the captured camera image (blue squares). After two internal simulations, the position of the stimulus is shown with red crosses with a mean error of 35.72 pixels (red disk), which means a 1.3% of the total image size and 3° of robot's motion.

The error coming from the saccadic movements are chiefly due to the visual fovealization since it reduces the information available around the image periphery, which causes precision loss in determining the initial stimulus location.

3.1.5.2. Execution

It is known that two components of a motoneuronal control signal generate the saccadic eye movement in humans (Bahill et al., 1975). These components correspond to an initial saccade and a corrective saccade.

Based on this fact, after realizing the sequence of moves suggested by the model (initial or approaching saccade), a second tuning or corrective motion is executed (see **Figure 10**). These two execution steps are described next.

- *Approaching saccadic movement:* the current image of the camera, S_t, and the target image S_{t+1}, with the stimulus at the center, are fed into the system. These inputs go through SOIMA and generate a suggested motor command M_t^* that is applied to

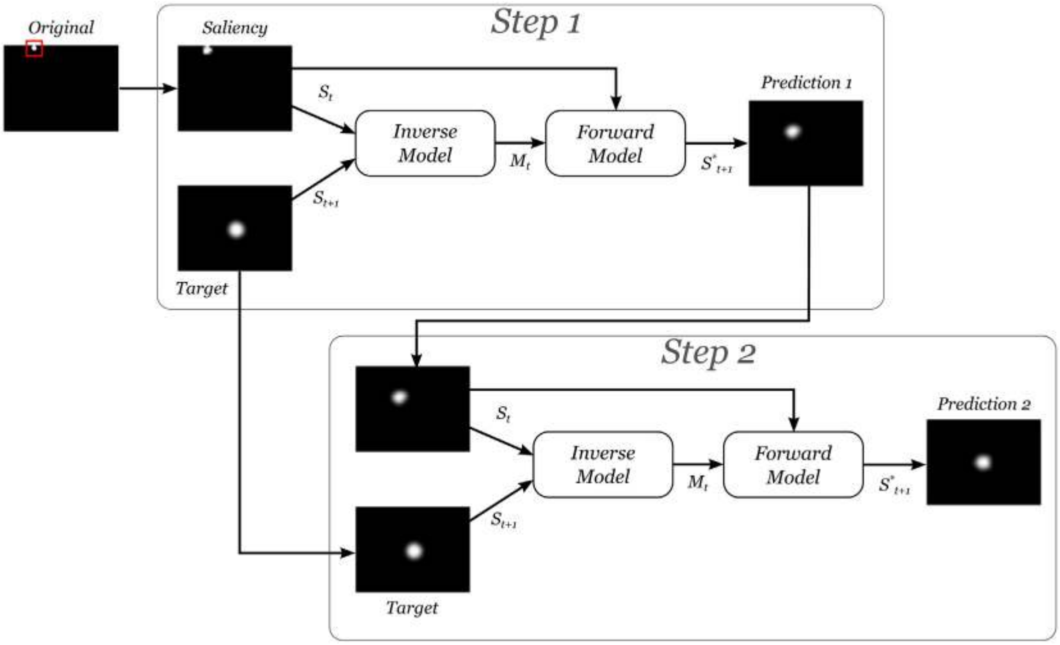

FIGURE 8 | **Typical prediction example**.

the Nao Robot. The error of the resulting stimulus location in the picture is calculated in order to know whether a second movement, for better accuracy, is necessary.

- *Tuning saccadic movement:* if the error is greater than a threshold of 10 pixels (0.1% of image surface and 0.8° of robot's movement), the actual image is fed back again to SOIMA to obtain an additional tuning motor command in order to reach the desired position accurately.

As opposed to the control described in Section 3.1.5, in this case motor commands are actually executed in both movements. In practice, this means that both movements in this strategy could contain more than one simulation step.

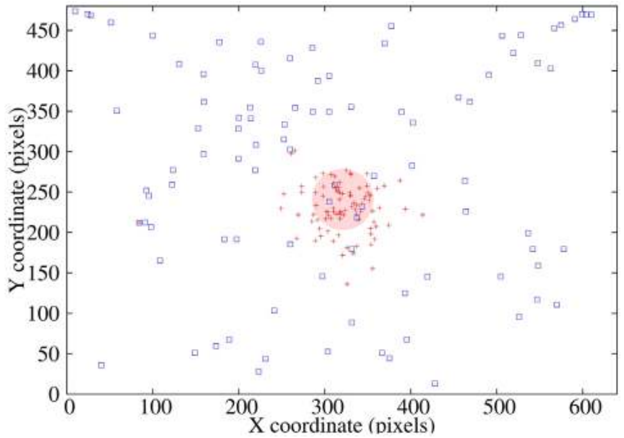

FIGURE 9 | Prediction performance: position of the stimulus after two predictions for 95 random initial positions.

The system was tested on 84 patterns and it was found an average error of 36.1 pixels on the original 640 × 480 pixels image, which corresponds to a 3.1° error on approaching saccadic movements. For the tuning motion, a mean error of 19.3 pixels was obtained corresponding to a 1.6° error (see **Figure 11**).

3.1.5.3. Tracking of Stimulus

In addition, we realized a tracking experiment. The stimulus was moved around the wall plane facing the robot, while the latter executed a centering or foveation task by means of the acquired saccadic control model. The purpose of this experiment is to show that even when the stimulus spatial reference with respect to the robot changes, and so do the perspective, the agent is able to effectively use the saccadic controller.

In **Figure 12,** we depict a path following task. Red arrows show the sequence of the path followed by the stimulus. Numbers associated with each arrow correspond to the saccadic movements executed to center the stimulus on the image. Finally, blue numbers and circles show the centering task error in pixels over every point of the path.

3.2. Hand–Eye Coordination

Coordination of visual perception with body movements is an important prerequisite for the development of complex motor abilities.

Visuomotor coordination refers to the process of mapping visuospatial information into patterns of muscular activation. Such mapping is learned through the interaction of the agent with

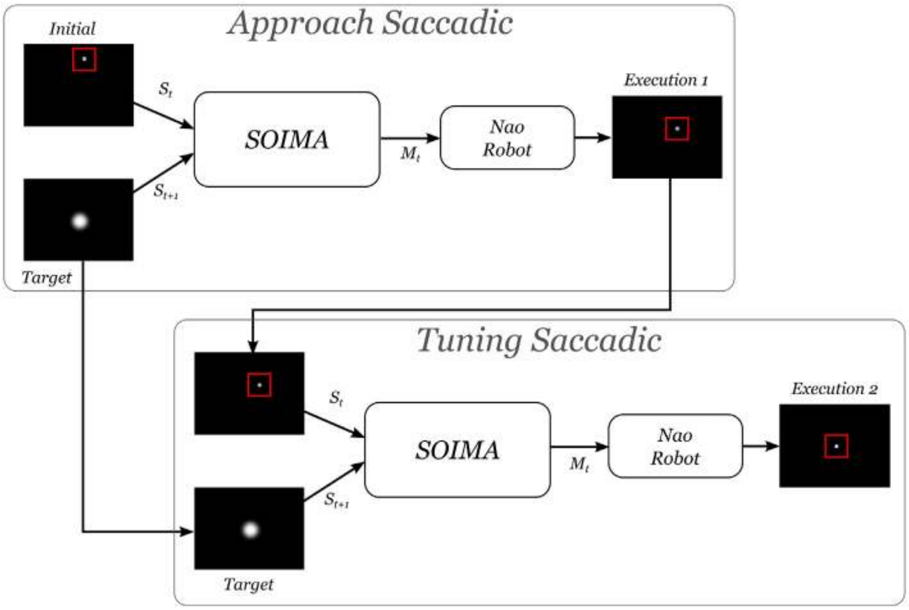

FIGURE 10 | Execution algorithm for saccadic movements. See text for explanation.

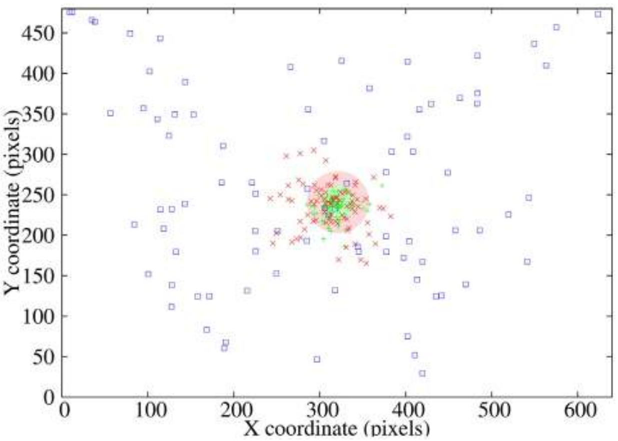

FIGURE 11 | Execution performance, including coarse and fine motions: initial stimulus locations in the picture through 84 test patterns (blue squares), ending stimulus locations during the approaching saccadic movement (red crosses), mean error (red circle), and final stimulus locations during the tuning saccadic movement (green crosses).

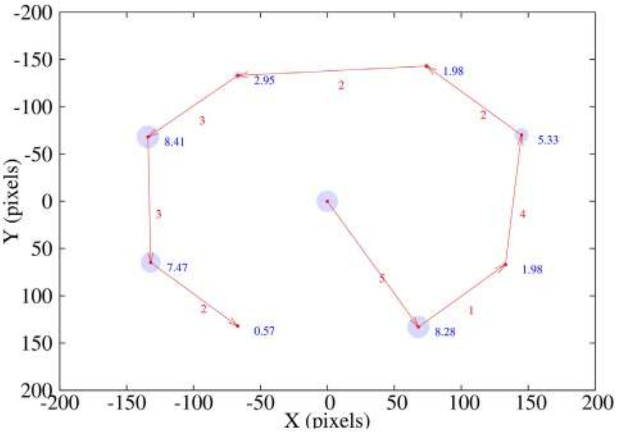

FIGURE 12 | Path followed by the stimulus on the robot's image plane. Red arrows and numbers show the path sequence and the number of saccadic movements in each centering task. Blue circles and numbers depict the error in pixels after centering the stimulus.

his environment. In particular Hand–Eye Coordination (*HEC*) refers to the coordinated use of the eyes with one or both hands to perform a task.

Here, we propose an implementation of the SOIMA for the learning of a sensory–motor scheme that allows *HEC* in a NAO humanoid robot. Once this coordination is learned then, given a particular posture (i.e., arm and hand postures) of the robot, the system should provide a head posture such that the hand appears in the visual field. It is worth mentioning that a posture is determined by absolute joint angles.

The SOIMA structure proposed can be seen in **Figure 13**, showing the integration of a *HEC* Multi-Modal Representation, together with the saccadic control presented in the previous section.

Here, *V* is an 80 × 80 SOM coding the coordinates of the position of the robot's hand in the image plane. The image was obtained from the lower camera in the head of the robot. The coordinates were estimated with the use of the ARToolkit library,[6] using a fiduciary marker in the hand of the robot.

Head is a 100 × 100 SOM coding the values of the two degrees of freedom of the head *(yaw and pitch)*. *Arm* is a 80 × 80 SOM coding *shoulder pitch, shoulder roll, elbow yaw, and elbow roll* of the left robot arm. Finally, MMR_{eh} is a 150 × 150 SOM that codes the Multi-modal Representation of the sensory–motor scheme. The SOMs were trained using 6453 random patterns. The saccadic control used the same SOMs described in the previous experiment.

In this experiment, the system for saccadic control was used only as a tool for the training and testing of the *HEC* system. Given that both the *HEC* and the saccadic control system use the same visual input, the map *V* was the same as previously defined.

The experiments were carried out using the simulated NAO in an empty arena as shown in **Figure 14**.

Motor babbling was used in order to train the Hebbian associations for the *HEC* system. The general procedure for training was:

- Execution of random head and arm movements.
- Verify whether the marker was detected. When detected, the positions of head and arm as well as the visual information were fed into the system.
- The connection between the winning units in each of the three maps was reinforced.

During training, the saccadic control system was used to increase the variability and precision of the patterns used for the Hebbian associations. In the cases where the marker was found in the image, but not in the fovea, the saccadic control system was used to center the stimulus.

Two tests were carried out to assess the full system:

- First, the robot performed random arm movements and these were fed into the system; the head would then follow successfully the hand position, centering the stimulus in the foveal area.
- Second, a pointing test was implemented. A marker was randomly placed in the arena, the system would then perform a random exploration of the visual space. Once the maker was in sight, the saccadic control system would center the stimulus. The *HEC* system would then successfully position the hand before the marker.

Video material on both tests is available at the following url links: Test 1 on simulated robot[7]; Test 1 on real robot[8]; Test 2 on simulated robot[9]; Test 2 on real robot.[10]

[6] www.hitl.washington.edu/artoolkit

[7] https://youtu.be/agrkeUxiQZA

[8] https://youtu.be/7h_luKEre5s

[9] https://www.youtube.com/watch?v=BrFS7EWz4kc

[10] https://www.youtube.com/watch?v=ed7WkMgjybo

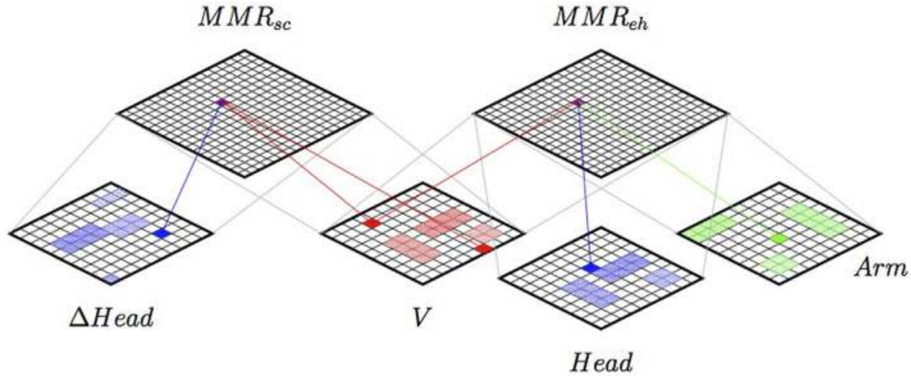

FIGURE 13 | The SOIMA up-scaled when implementing a Hand–Eye Coordination strategy, including the saccadic control presented in the previous section. Two Multi-Modal Representations (sensory–motor schemes) are coupled together through modal maps. See explanation in text.

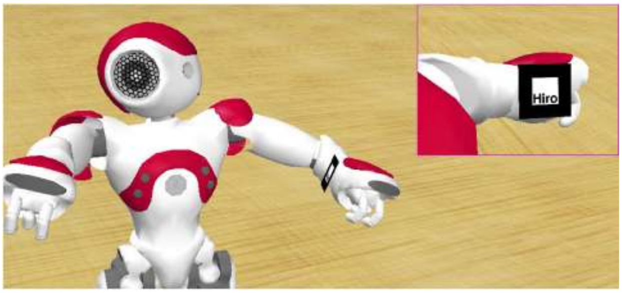

FIGURE 14 | Hand–Eye Coordination task experimental setup. Given particular arm and hand postures of the robot, the system provides a head posture such that the hand appears in the visual field.

4. DISCUSSION AND CONCLUSION

The relevance of modeling sensory–motor schemes relies on the fact that they are considered to be the fundamental unit of analysis for cognitive processes and skills under the cognitive robotics school of thought (Lungarella et al., 2003). Cognition relies on self-organized structures integrating sensory–motor information. As internal models create naturally a multi-modal representation of sensory–motor flows, they have been extensively studied as a suitable mechanism for sensory–motor integration.

Based on these considerations, we developed a new computational architecture called SOIMA, drawing biological inspiration from the theory of convergence–divergence zones of the cerebral cortex proposed by Damasio (1989) and Meyer and Damasio (2009) and from the self-organizing properties of the brain.

In order to introduce the architecture and prove its feasibility and performance, we implemented two case studies. The first experiment implemented a strategy of saccadic movement control consisting in centering a salient stimulus in the visual sensory space using the SOIMA. This experiment shown that the SOIMA approach allows to cope with vision issues regarding the input space dimensionality. The second case study implemented a Hand–Eye Coordination strategy allowing to show how the SOIMA can be scaled upwardly in order to model more complex cognitive tasks.

The SOIMA integrates important qualities of online learning and introduces a novel form of internal models implementation not reported before. Even though there exists work showing coupled Self-Organized Networks (Hikita et al., 2008; Luciw and Weng, 2010; Morse et al., 2010b; Lallee and Dominey, 2013), our proposal goes a step further in that we model the predictive capabilities of the human cognitive machinery by means of internal models. However, current internal models architectures show major drawbacks so as to model cognition under the embodiment hypothesis constraints (e.g., independent coding of the inverse and forward models). The main attributes of the SOIMA provide means for autonomous sensory–motor integration, as it allows for multi-modal activation patterns to organize themselves into a coherent structure through Hebbian association, creating thus a multi-modal grounded representation. The bi-directional capability of the SOIMA allows this representation to become a sensory–motor scheme available as both a forward (predictive) and an inverse (controlling) model. The lack of this property is precisely one of the main limitations of current cognitive architectures. This bi-directional mechanism provides, thus, a unified substrate allowing for a true sensory–motor integration space and for coherent sensory–motor learning strategies.

We would like to highlight five main features making the SOIMA stand apart from current implementations of internal models and sensory–motor schemes. The first three features have been tested in the case studies presented here, the remaining two represent current work:

- *Modularity and scalability*: the experiments reported here exemplify the modular character of SOIMA. This feature redounds in an integrated learning strategy and allows for the scalability of the system. The architecture is modular in that the logical structure of the *MMR* is not hardwired, but develops as the agent interacts with the environment. This in turn provides means for the construction of sensory–motor schemes that can be sequentially re-enacted to accomplish a particular task. The workings of the architecture enable the system to learn online both, the FM and the IM, in an integrated way. New examples of sensory–motor schemes are acquired as the agent experiences the world; thus, incorporating new knowledge for later use. A first example of the scalability of the system was reported here. Every new sensory–motor scheme generates a new *MMR*, coding a particular IM–FM coupling. Thus, different sensory–motor schemes can be coupled together in order to increase the sensory–motor capabilities of the agent. In this sense, the system is scalable. In summary, the SOIMA should be seen as a core unit for building more complex structures allowing to re-use previous knowledge.

- *Bidirectionality*: given the internal coding of the relations between the sensory and motor modalities, the SOIMA works as either a Forward or an Inverse model. The connections between the maps admit the bidirectional flow of information, therefore choosing the model depends on the question asked and the available input. As mentioned before, current implementations use mostly MLP networks, with one or various networks coding the Forward model and separately networks coding the Inverse model, which obligates to use different learning strategies for each model. Our approach synthesizes the coupled model on the same substrate, conferring connectivity enhancements that allow for integrated learning strategies and sensory–motor scheme maps.
- *Temporality*: different moments in time for the sensory situation are coded in the same map. The temporal relations between situations are coded in the Hebbian connections between maps. As a consequence, several time steps can be integrated into the same sensory–motor scheme. Another concern regarding temporality is stability. That is, whether the system will be able to cope with environmental changing conditions, i.e., become stable after some perturbation. This is indeed a major concern that is currently being investigated. Future work includes experimenting with other kinds of self-organizing maps (e.g., Dynamic SOMs). It is expected that these other maps would also allow for sensory–motor scheme reconfiguration in the long run, if ever the available knowledge is not enough to properly model contingent task changes.
- *Motor Mapping*: the characterization of the information in the motor map should allow for trajectory generation in the motor space. Moving in the motor map from unit to unit would then have a mapping in the physical space of the agent.
- *Robustness to lack of information*: the structure proposed can be the base for capabilities such as action recognition. The agent codes a SOIMA based on its own experience and its own sensory–motor model; however, when observing the execution of an action some information would not be available (i.e., proprioceptive information). The lack of this information should not represent a problem as the activation produced by the available input should propagate the activation in the rest of the architecture.

The results presented here are encouraging and permit us to assert that the organization and functioning of the SOIMA is promising for undertaking research in further directions. In the context of Grounded Cognition, we consider that our work constitutes a biologically plausible computational approach, effective for the development of complex cognitive behavior models. As such, we hope that the SOIMA concept will enable the study and test of diverse hypotheses on the underpinning processes of cognition and the development of artificial agents exhibiting coherent behavior in their environment.

AUTHOR CONTRIBUTIONS

EE, JH, and BL designed research. EE performed research. EE, GS, JH, and BL wrote the paper.

ACKNOWLEDGMENTS

The work from GS has been funded from the European Unions Seventh Framework Programme (FP7/2007-2013) under grant agreement n. 609465, related to the EARS (Embodied Audition for RobotS) project.

REFERENCES

Arceo, D. C., Escobar, E., Hermosillo, J., and Lara, B. (2013). Model of a mirror neuron system in an artificial autonomous agent. *Nova Scientia* 5, 51–72.

Bahill, A. T., Clark, M. R., and Stark, L. (1975). Glissades – eye movements generated by mismatched components of the saccadic motoneuronal control signal. *Math. Biosci.* 26, 303–318. doi:10.1016/0025-5564(75)90018-8

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 1177–1187. doi:10.1098/rstb.2003.1319

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639

Blakemore, S.-J., Wolpert, D., and Frith, C. (2000). Why can't you tickle yourself? *Neuroreport* 11, R11–R16. doi:10.1097/00001756-200008030-00002

Cang, J., and Feldheim, D. A. (2013). Developmental mechanisms of topographic map formation and alignment. *Annu. Rev. Neurosci.* 36, 51–77. doi:10.1146/annurev-neuro-062012-170341

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Comput.* 1, 123–132. doi:10.1162/neco.1989.1.1.123

Dearden, A. (2008). *Developmental Learning of Internal Models for Robotics*. Ph.D. thesis, Imperial College London, London.

Demiris, Y., and Khadhouri, B. (2006). Hierarchical attentive multiple models for execution and recognition of actions. *Rob. Auton. Syst.* 54, 361–369. doi:10.1016/j.robot.2006.02.003

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behav. Brain Sci.* 27, 377–396. doi:10.1017/S0140525X04000093

Hebb, D. (1949). *The Organization of Behavior A Neuropsychological Theory*. New York: Lawrence Erlbaum Associates.

Hikita, M., Fuke, S., Ogino, M., Minato, T., and Asada, M. (2008). "Visual attention by saliency leads cross-modal body representation," in *Development and Learning, 2008. ICDL 2008. 7th IEEE International Conference on* (Monterey, CA: IEEE), 157–162.

Hoffmann, M., Marques, H., Hernandez Arieta, A., Sumioka, H., Lungarella, M., and Pfeifer, R. (2010). Body schema in robotics: a review. *IEEE Trans. Auton. Ment. Dev.* 2, 304–324. doi:10.1109/TAMD.2010.2086454

Jordan, M. I., and Rumelhart, D. E. (1992). Forward models: supervised learning with a distal teacher. *Cogn. Sci.* 16, 307–354. doi:10.1207/s15516709cog1603_1

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Res. Bull.* 44, 107–112. doi:10.1016/S0361-9230(97)00094-4

Kaiser, A., Schenck, W., and Möller, R. (2013). Solving the correspondence problem in stereo vision by internal simulation. *Adapt. Behav.* 21, 239–250. doi:10.1177/1059712313488425

Karaoguz, C., Dunn, M., Rodemann, T., and Goerick, C. (2009). "Online adaptation of gaze fixation for a stereo-vergence system with foveated vision," in *Advanced Robotics, 2009. ICAR 2009. International Conference on* (Munich: IEEE), 1–6.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* 9, 718–727. doi:10.1016/S0959-4388(99)00028-8

Kohonen, T. (1990). The self-organizing map. *Proc. IEEE* 78, 1464–1480. doi:10.1109/5.58325

Lallee, S., and Dominey, P. F. (2013). Multi-modal convergence maps: from body schema and self-representation to mental imagery. *Adapt. Behav.* 21, 274–285. doi:10.1177/1059712313488423

Lara, B., and Rendon-Mancha, J. M. (2006). "Prediction of undesired situations based on multi-modal representations," in *Electronics, Robotics and Automotive Mechanics Conference, 2006*, Vol. 1 (Cuernavaca: IEEE), 131–136.

Leigh, R. J., and Zee, D. S. (1999). *The Neurology of Eye Movements*, 3rd Edn. Oxford: Oxford University Press.

Li, P., Zhao, X., and Mac Whinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cogn. Sci.* 31, 581–612. doi:10.1080/15326900701399905

Luciw, M., and Weng, J. (2010). Top-down connections in self-organizing hebbian networks: topographic class grouping. *IEEE Trans. Auton. Ment. Dev.* 2, 248–261. doi:10.1109/TAMD.2010.2072150

Lungarella, M., Metta, G., Pfeifer, R., and Sandini, G. (2003). Developmental robotics: a survey. *Connect. Sci.* 15, 151–190. doi:10.1080/09540090310001655110

Maravita, A., and Iriki, A. (2004). Tools for the body (schema). *Trends Cogn. Sci.* 8, 79–86. doi:10.1016/j.tics.2003.12.008

Maravita, A., Spence, C., and Driver, J. (2003). Multisensory integration and the body schema: close to hand and within reach. *Curr. Biol.* 13, R531–R539. doi:10.1016/S0960-9822(03)00449-4

Mayor, J., and Plunkett, K. (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychol. Rev.* 117, 1. doi:10.1037/a0018130

Meyer, K., and Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends Neurosci.* 32, 376–382. doi:10.1016/j.tins.2009.04.002

Möller, R., and Schenck, W. (2008). Bootstrapping cognition from behavior – a computerized thought experiment. *Cogn. Sci.* 32, 504–542. doi:10.1080/03640210802035241

Morse, A. F., Belpaeme, T., Cangelosi, A., and Smith, L. B. (2010a). "Thinking with your body: modelling spatial biases in categorization using a real humanoid robot," in *Proc. of 2010 Annual Meeting of the Cognitive Science Society* (Portland: Cognitive Science Society), 1362–1368. Available at: http://www.proceedings.com/09137.html

Morse, A. F., De Greeff, J., Belpaeme, T., and Cangelosi, A. (2010b). Epigenetic robotics architecture (era). *IEEE Trans. Auton. Ment. Dev.* 2, 325–339. doi:10.1109/TAMD.2010.2087020

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., and Spivey, M. J. (2013a). Computational grounded cognition: a new alliance between grounded cognition and computational modeling. *Front. Psychol.* 3:612. doi:10.3389/fpsyg.2012.00612

Pezzulo, G., Candidi, M., Dindo, H., and Barca, L. (2013b). Action simulation in the human brain: twelve questions. *New Ideas Psychol.* 31, 270–290. doi:10.1016/j.newideapsych.2013.01.004

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., Spivey, M., and McRae, K. (2011). The mechanics of embodiment: a dialogue on embodiment and computational modeling. *Front. Psychol.* 2:5. doi:10.3389/fpsyg.2011.00005

Pfeifer, R., and Bongard, J. (2007). *How the Body Shapes the Way We Think: A New View of Intelligence*. Cambridge: MIT press.

Pfeifer, R., and Scheier, C. (2001). *Understanding Intelligence*. Cambridge: MIT press.

Rochat, P. (1998). Self-perception and action in infancy. *Exp. Brain Res.* 123, 102–109. doi:10.1007/s002210050550

Rochat, P., and Morgan, R. (1998). Two functional orientations of self-exploration in infancy. *Br. J. Dev. Psychol.* 16, 139–154. doi:10.1111/j.2044-835X.1998.tb00914.x

Schenck, W. (2008). *Adaptive Internal Models for Motor Control and Visual Prediction. Number 20*. Berlin: Logos Verlag Berlin GmbH.

Schenck, W., Hoffmann, H., and Möller, R. (2011). Grasping of extrafoveal targets: a robotic model. *New Ideas Psychol.* 29, 235–259. doi:10.1016/j.newideapsych.2009.07.005

Schillaci, G., Hafner, V. V., and Lara, B. (2012a). "Coupled inverse-forward models for action execution leading to tool-use in a humanoid robot," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction* (Boston, MA: ACM), 231–232.

Schillaci, G., Lara, B., and Hafner, V. (2012b). "Internal simulations for behaviour selection and recognition," in *Human Behavior Understanding, Volume 7559 of Lecture Notes in Computer Science* (Berlin: Springer), 148–160.

Schillaci, G., Hafner, V. V., Lara, B., and Grosjean, M. (2013). "Is that me?: sensorimotor learning and self-other distinction in robotics," in *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction* (Tokyo: IEEE Press), 223–224.

Svensson, H., and Ziemke, T. (2004). "Making sense of embodiment: simulation theories and the sharing of neural circuitry between sensorimotor and cognitive processes," in *Presented at the 26th Annual Cognitive Science Society Conference* (Chicago, IL).

Udin, S. B., and Fawcett, J. W. (1988). Formation of topographic maps. *Annu. Rev. Neurosci.* 11, 289–327. doi:10.1146/annurev.ne.11.030188.001445

Westerman, G., and Miranda, E. R. (2002). Modelling the development of mirror neurons for auditory-motor integration. *J. New Music Res.* 31, 367–375. doi:10.1076/jnmr.31.4.367.14166

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636. doi:10.3758/BF03196322

Wolpert, D. M., Ghahramani, Z., and Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends Cogn. Sci.* 5, 487–494. doi:10.1016/S1364-6613(00)01773-3

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882.

Wolpert, D. M., and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Netw.* 11, 1317–1329. doi:10.1016/S0893-6080(98)00066-5

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.