Restricted Boltzmann Machine with Transformation Units for Laser Image Processing in a Mirror Neuron System Architecture

Junpei Zhong, Cornelius Weber and Stefan Wermter
Department of Computer Science,
University of Hamburg,
Vogt-Koelln-Str. 30,
22527 Hamburg, Germany
Email: {zhong,weber,wermter}@informatik.uni-hamburg.de

***Abstract*—In the mirror neuron system, the canonical neurons play a role in object shape and observer-object relation recognition. However, there are almost no functional models of canonical neurons towards the integration of these two functions. We attempt to distributed represent the relative position between the object and the robot in a neural network model. Although at present some generative models based on the Restricted Boltzmann Machine can code the image transformation in continuous images, what we need to accomplish in canonical neuron modeling is different from the requirements of modeling transformation in video frames. As a result, we propose a novel model called "Restricted Boltzmann Machine with Transformation Units", which can represent the relative object positions based on laser images. The laser sensor provides binary and accurate images and can further be connected with other models to construct a unified architecture of the mirror neuron system.**

I. Introduction

Since Rizzolatti and his colleagues found that some neurons in the F5 area of macaque monkeys' premotor cortex fired when the monkeys did actions like reaching for something or biting a peanut, the so-called mirror neurons have become a significant research topic that explains many social behaviors of human beings. A number of computational models ([1], [2], [3], [4], [5], [6], [7], [8]) have been designed to model different functions of the mirror neuron system.

Moreover, a lot of research points out that the object affordance, the relative position between an observer and an object, should be considered in an integral mirror neuron system model, because actions cannot be understood without understanding object affordances (e.g. [9], [**?**]). This function may be realized by the canonical neurons which are active when the object that can be grasped by movements is observed, as if the brain is foreseeing a possible interaction with this object and preparing itself accordingly. These neurons act with a mechanism of recognizing the object affordance with visual or other stimuli [10]. This does not lead to the motor action itself, but to the semantic knowledge about the actions. These kinds of canonical neurons also exist in the ventral premotor area F5 [11], [12].

As a result, if we judge the property of action understanding with consideration of the object affordance of the whole mirror neuron system, a proper computational model is necessary to consider both the object shape and the observer-object relation. For instance, in [13], [14] and [15] of the MNS and MNS2 model, the authors solved this problem by manually calculating it in a geometric way.

In this paper, we propose an ongoing model which emphasizes the representation of the relative position of the object. Specifically, this position information is represented in a distributed manner in units called Transformation Units.

In the next section, we introduce the main architecture of the Restricted Boltzmann Machine and its modified version of coding object transformations. In section 3, the Restricted Boltzmann Machine with transformation units is presented. Then the experiment of the novel Restricted Boltzmann Machine is described in section 4. At the end we close with discussions and conclusions.

II. Restricted Boltzmann Machine and Related Works

In the neural networks community, the problem of representing relative positions can be simply converted into a similar problem of coding the object transformation in a distributed representation of a network. In this way the observer-object relation problem can be considered as a modified version of the transformation problem if we fix the observer at the origin and regard the image transformation as a representation of the observer-object relation.

A. Restricted Boltzmann Machine

A binary Restricted Boltzmann Machine (RBM) consists of a layer of visible units $v \in \{0, 1\}$ and hidden units $h \in \{0, 1\}$. The connections between the hidden units and the visible units are symmetric. Without interconnections between the visible units or the hidden units, the hidden units are conditionally independent. The probability distribution over the hidden units and the visible units is defined by

$$P(v, h) = P(V = v, H = h) = exp(v^T b + h^T c + v^T W h)/Z \quad (1)$$

where b is the vector of biases for the visible units, c is the vector of biases for the hidden units, W is the matrix of connection weights, and $Z(v,h) = \sum_{v,h} exp(v^T b + h^T c + v^T W h)$.

The RBM's definition ensures that all the distributions of values in hidden units $P(h|v)$ are only dependent on the values of visible units; the values in visible units are determined in the same way.

The updating rule for parameters W is:

$$\Delta W = \langle vh \rangle_{p(h|v,W)} - \langle vh \rangle_{p(v,h|w)} \tag{2}$$

the notation $\langle \cdot \rangle_{P()}$ is the expectation over the probability distribution $P()$.

B. RBM for Image Transformation

The first related paper about this topic of image transformation was advocated by Memisevic and Hinton[16]. They established a gated Restricted Boltzmann Machine which is able to extract distributed, domain-specific representations of image patch transformation. The model developed a mechanism to represent domain-specific motion features in the hidden units[1]. It works in two pathways: in the gated regression, from a given pair of two observed images, the hidden units are coded as a transformation; while in the modulated filters pathway, given the codes for the transformations as latent variables, the model can subsequently predict transformations of the next successive image. However, due to its tensor parameterization, the computational load of the training is quite high. Later Memisevic and Hinton developed a factored version of GRBM that uses three low-rank matrices as a factor instead of three-way parameters [17] to reduce the computational effort.

Based on the older version of GRBM, Taylor et al. proposed another method to reduce the complexity [18] by adding the convolution with a compact parameterization so that the invariance of relative transformation does not need to be re-learned.

A temporal Restricted Boltzmann Machine [19], [20] can be used as well to model sequences where the decision at each step requires some context information from the past. It contains one RBM model in each time step and these models are connected with latent to latent connections between the time steps. Therefore, it can also be used as a kind of transformation representation when the previous image provides a type of "temporal bias" to the hidden variables. Recently, Sutskever et al. [21] also proposed an improved type of TRBM by adding recurrent connections between the hidden units and visible units.

Recent research by Hinton [22] introduces the concept of "local capsules" which represents different kinds of transformations (poses, angles, lightings) so that it is possible to encapsulate the results with the inputs. The basic form of this model is an explicit representation of the pose. Furthermore, more levels of these capsules will be able to model different transformations.

[1]These hidden units are not exactly the same as hidden units in generic RBM, but a kind of latent variables.

III. Restricted Boltzmann Machine with Transformation Units

In this section we propose the Restricted Boltzmann Machine with Transformation Units (RBM-TU). As part of the mirror neuron system model, we expect the information in the transformation units of the object's relative position can be further linked with PB units of the Simple Recurrent Network with Parametric Bias (SRN-PB) [23] so that these two units can interact as in Fig. 1. The RBM-TU is a modified version of the RBM. It has the same architecture as RBM except the full connections between transformation units and hidden layer (Fig. 2), this network can be regarded as a hybrid architecture consisting of a generative model and a discriminative model. The RBM recognizes and reconstructs the learned images, while the transformation units represent the relation between the objects and the sensor.

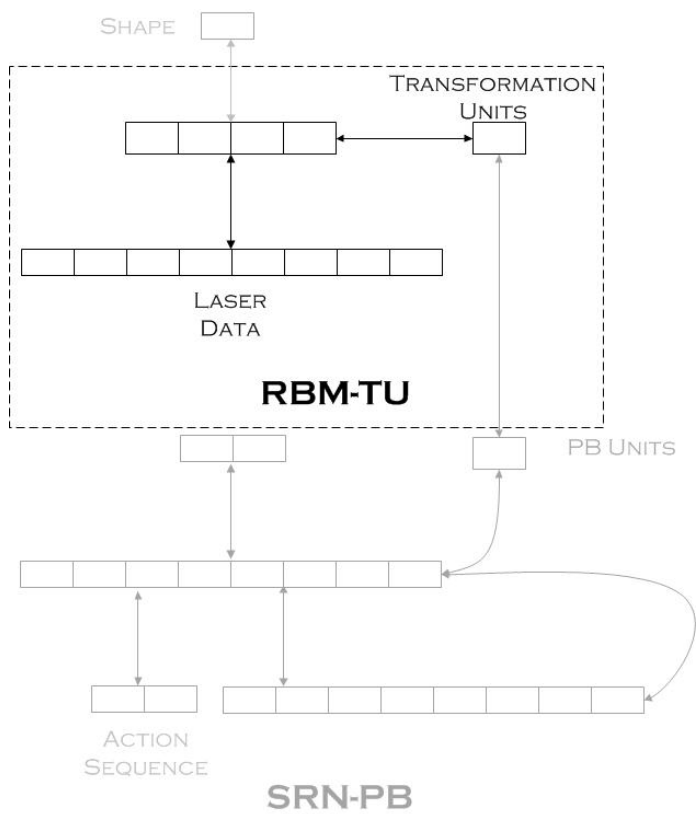


Fig. 1. The proposed computational model of a mirror neuron system towards action understanding: the RBM-TU represents information on the object shape and position, and the SRN-PB processes the action sequence together with the parametric biases units so that the structure of the input sequence can be characterized and recognized. The network within the dashed rectangle is the one we are proposing here.

Training is done by placing an object within the laser sensor area while the robot and the object stay still. The algorithm for the training is introduced in the next section. After training, the property of transformation units enables us to recognize

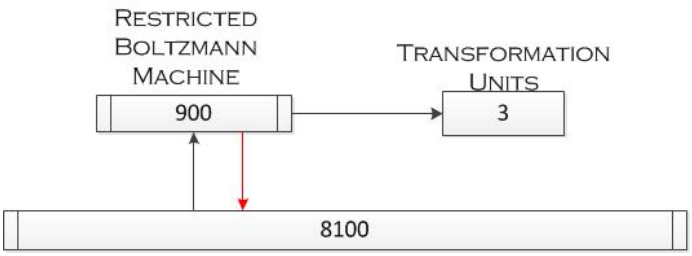


Fig. 2. Restricted Boltzmann Machine with transformation units Model: the laser information input into the visible units as a $90 * 90$ image. The hidden layer contains 900 units, which are fully connected with three transformation units. We simply use a linear function in the transformation units.

the transformation of the input compared to the training sets. Comparing our approach to the GRBM, it is more straightforward to implement and reduce the computational effort substantially because the updating of the RBM weights and the transformation units' weights are separate algorithms.

A. RBM Training

In the training mode, during each epoch, the weights between hidden units and visible units are updated in the same way as the generic RBM model by Contrastive Divergence. After that, the output values of the hidden units with the sigmoid function become inputs to the transformation units. In this way, the hidden units and the transformation units form an independent two-layer network, in which the first layer consists of the hidden units and the second layer consists of the transformation units. The connection weights between the hidden units and the transformation units are updated by back-propagated the error from the transformation units. For the position of the object during training, which acts as a reference position, the target values are set to zero. The activation function of hidden units can be expressed as:

$$H_h = \frac{1}{1 + e^{(-s_h)}} \tag{3}$$

where s_h is the internal value of the hidden layer, and H_h is the output value of the hth hidden unit.

The activation function of transformation units is a linear function:

$$TU_i = \sum_h H_h W_{h,i} \tag{4}$$

where W is the connection weight matrix between hidden units and transformation units. The updating rule of connection weights between transformation units and hidden units is:

$$\Delta W_{h,i} = \eta \delta_i H_h \tag{5}$$

where η is the learning rate of transformation weights, and δ is the output error for the transformation units. The algorithm for a complete epoch can be depicted as follows:

[2]The word "batch" here means dividing the training set into small sets, which is very advantageous for parallel computing on GPU boards or in Matlab [24]. When obtaining each patch of training data, the object remains in the origin position, and the laser range sensor scans the whole area 100 times.

Algorithm 1 RBM-TU Training

Extract a mini-batch of data sets from the training set.[2]
Update the connection weights between the hidden units and the visible units using Contrastive Divergence.
Calculate the values of hidden units given the updated weights and input.
Update the connection weights between the hidden units and the transformation units by Eq. 5.

In the recognition mode, the connection weights are fixed and the values in the transformation units are calculated with laser images of different relative positions of the object. We expect that the difference of dense coding in the hidden layer due to the relative position will lead to variations of the transformation units. In the next section, an experiment of representing untrained positions will be conducted to examine the plausibility of this model.

In the experiment, an Aldebaran NAO [25] is equipped with a laser head (Fig. 3). The laser head is built based on a URG-04LX laser sensor by Hokuyo [26], with the angle coverage of 240 degrees. The reason why we use a laser sensor is that understanding the object distance often needs calibration of vision processing, which takes great efforts in the robot perception process and needs to be re-calibrated in different environments. However, if we consider the perception development of human beings, the tactile sensation plays an important role in order to modulate the vision reference of a human [**?**]. Therefore, in this experiment, we attempt to simulate the tactile sensing via the laser sensor and use its information to represent the information on the object position, which is a more accurate position information than by only using a vision system. Further experiments may use it as an automatically calibration tool of robot vision system.

IV. Experiment Descriptions

A. Experiment Setup

Our environment is a normal office and an object is placed in front of the NAO for the training. The range sensor produces 500 images as data sets. These images have small variations because of angle variations of the laser beams and noise of the sensor, although we keep the positions of robot and object constant (Fig. 4). The two-dimensional dotted images of the laser head's surrounding are captured and learned (Fig. 6). During the RBM training, we select 100 images as one mini-batch to update the weights once after the whole mini-batch learning is finished. As soon as the training is over, the same object is placed in different positions from the trained position as long as they can be scanned by the laser sensor (the coordinates of these 20 example positions are depicted in Fig. 5).

The laser sensor in this experiment can be considered as a "measurement" method to survey the surrounding objects, and each part of the robot remains in the same position during

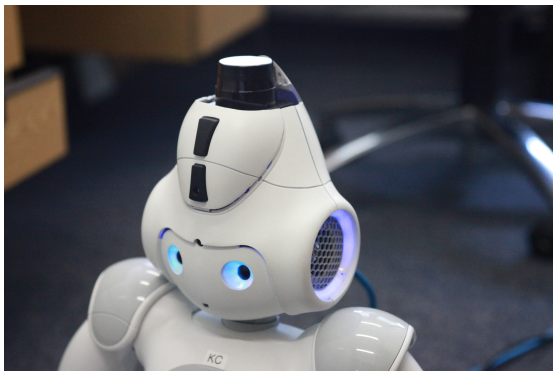
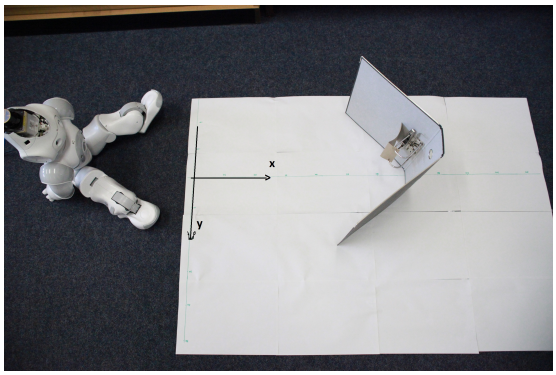


Fig. 3. Laser sensor in the NAO head



(a) Top view of relation between the NAO and the object: the axes show the reference coordinates for the objects and the robot.

(b) Side view of relation between the NAO and the object

Fig. 4. The NAO (left) senses different positions of the object (right, a folder) through the laser sensor. Markers on the ground are used to indicate the relative position between the robot and the object.

training and recognition so that the transformation of the objects equals the relative positions of the objects with respect to the robot.

Binary images obtained from the laser sensor contain the information on whether the laser beam encounters an object and on the distance between sensor and object. These images are processed and reconstructed in the Restricted Boltzmann Machine. At the same time, we examine the relation between the representations in the transformation units and the actual transformation.

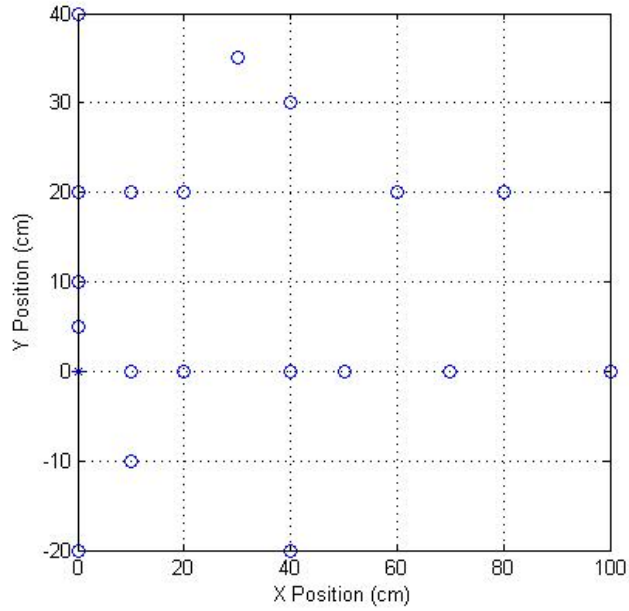


Fig. 5. Locations of 20 sample points: the object remains in the same position $(0, 0)$ during training (star mark). Then in recognition, 20 different sample points are picked to locate the object in order to test the values in transformation units (circle marks). The spine of the folder is always along with x axis.

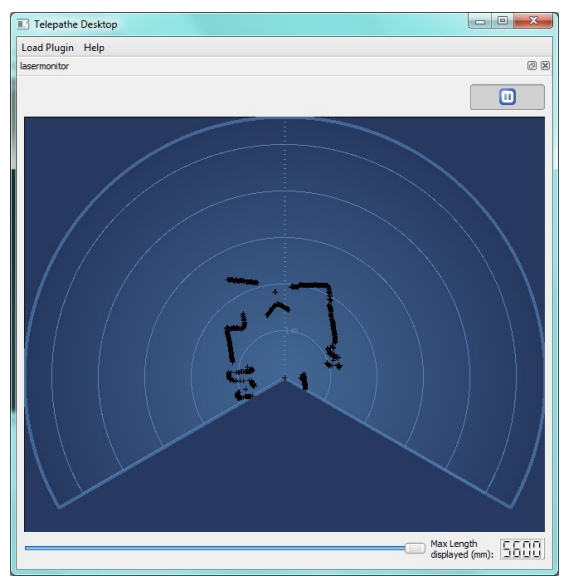
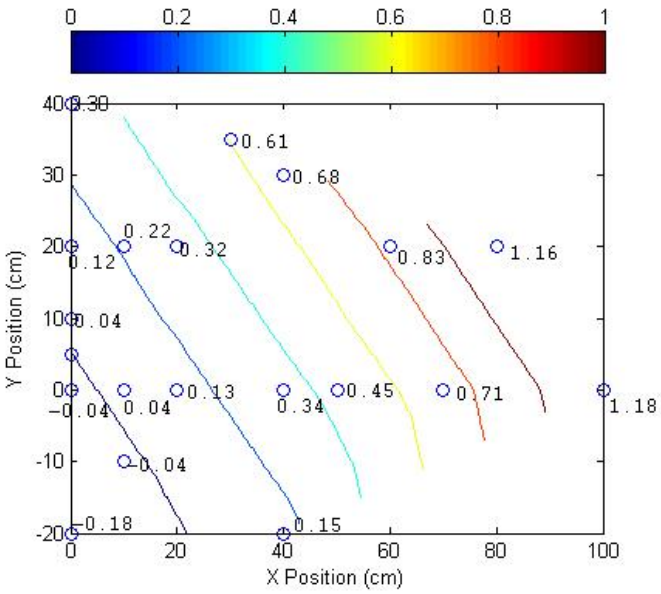
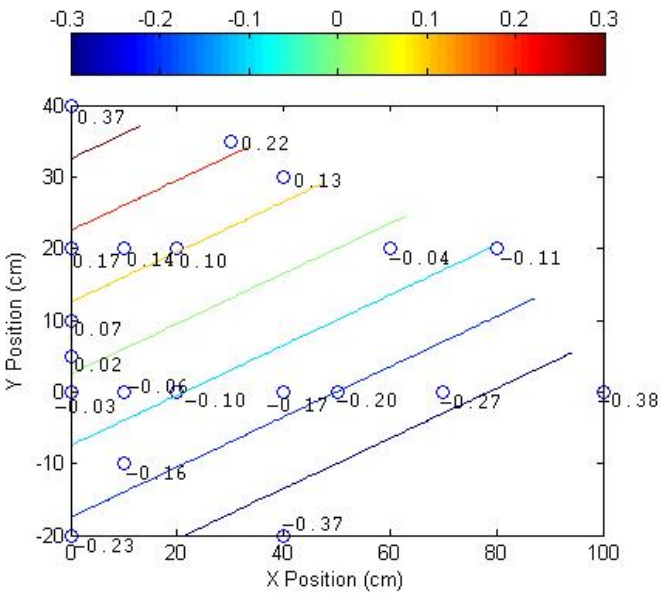


Fig. 6. Image obtained from NAO laser head: the binary image is obtained through laser beams reflected by obstacles. By the calculation of the time difference between emitting and receiving we can obtain the distance of the obstacle which that beam encounters.

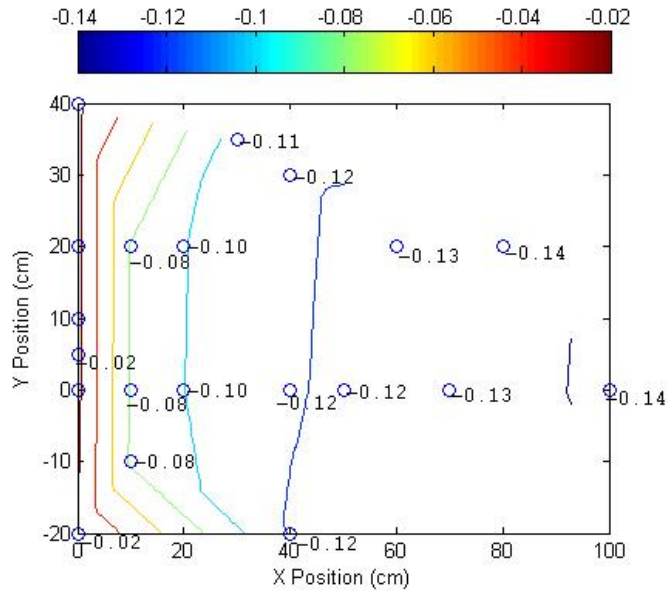


(a) Values of Transformation Unit 1



(b) Values of Transformation Unit 2

Fig. 7. Relation between Values in Transformation Units and Positions



(c) Values of Transformation Unit 3

Fig. 7. Relation between Transformation Values and Positions (cont.): The circles represent the values obtained from transformation units (values are accurate to 0.01; some values are omitted due to the limited figure space). The contour curves are obtained by linear interpolation. We can pinpoint the relative position of the object along the contours in these images by associating the values in the transformation units.

B. Experiment Results

We compare the values in transformation units with the selected untrained relative positions. Fig. 7 shows the relation between them.

Although we only select twenty positions as representatives within the range of the laser sensor, we can still identify the relation between transformation values and object positions. Fig. 7 uses interpolation to estimate the values in-between to better illustrate the transformation units. We can interpret this combination of the transformation units as the ability to locate a unique position of object; in this way we can regard the transformation units as a representation of the relative positions between the robot and the object.

V. Discussions and Conclusions

In this paper, we propose a novel hybrid architecture of a Restricted Boltzmann Machine with Transformation Units to model the functional model of observer-object relation representation in canonical neurons.

The RBM-TU model is a hybrid model which consists of a generative model (RBM) and a discriminative model (transformation units) to represent image transformation. Because of the independent learning processes between these two parts, the energy function is not as complex as for GRBM and other RBM-based image transformation models. Although it is not enough to apply as a video analysis for the transformation of consecutive images as GRBM does, it is straight-forward to be implemented because the independent RBM is the same as the original one.

The training of RBM and RBM-TU is almost the same as generic RBM, except that the updating of weights between hidden units and transformation units are back-propagated by the target value of zeros when the object is placed in the original position. Due to the architecture of the separate generative and discriminative models, the training of

RBM and transformation units are independent. This saves computational efforts compared to other RBM-based image transformation coding. The simple representation is suitable to process precise image representation, e.g. binary laser images.

Certainly, in terms of coding the object affordance like canonical neurons in the mirror neuron system, this computational model of the canonical neurons towards object affordance are not yet fully developed. An ideal functional model should function should represent both:

1) representing the shape of the object;
2) representing the relation of observer to the object.

The two requirements above can be related to the recognition and representation of "what" and "where" the object is. Although this problem is still far from solved, our experiment with laser sensing would become a reasonable foundational architecture of the multi-modal sensing to better perceive the information. Although we only use one object here to justify the transformation units, in further experiments RBM-TU with a soft-max label layer will be used to identify the object shapes as well.

Since the current experiment is done in the fixed context of room during both learning and recognizing the object position, In the next experiment we will attempt to combine the robot vision and laser sensor as multi-modal input so as to eliminate the room context and concentrate on the relative position of robot and object in the next experiment. Moreover, as presented in the first section, this model can be expected to be further integrated with SRNPB [23] to build up a complete mirror neuron system model.

ACKNOWLEDGMENT

This research has been partially supported by the EU projects RobotDoc under 235065 from the FP7, Marie Curie Action ITN and KSERA by the European Commission under n°2010-248085. We thank Katja Kösters for language suggestions.

REFERENCES

[1] Y. Demiris and B. Khadhouri, "Hierarchical attentive multiple models for execution and recognition of actions," *Robotics and Autonomous Systems*, vol. 54, no. 5, pp. 361–369, 2006.

[2] Y. Demiris and M. Johnson, "Distributed, predictive perception of actions: a biologically inspired robotics architecture for imitation and learning," *Connection Science*, vol. 15, no. 4, pp. 231–243, 2003.

[3] D. Wolpert and M. Kawato, "Multiple paired forward and inverse models for motor control," *Neural Networks*, vol. 11, no. 7-8, pp. 1317–1329, 1998.

[4] M. Haruno, D. Wolpert, and M. Kawato, "Hierarchical mosaic for movement generation," in *International Congress Series*, vol. 1250. Elsevier, 2003, pp. 575–590.

[5] J. Tani, M. Ito, and Y. Sugita, "Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using RNNPB," *Neural Networks*, vol. 17, no. 8-9, pp. 1273–1289, 2004.

[6] E. Borenstein and E. Ruppin, "The evolution of imitation and mirror neurons in adaptive agents," *Cognitive Systems Research*, vol. 6, no. 3, pp. 229–242, 2005.

[7] M. Elshaw, C. Weber, A. Zochios, and S. Wermter, "A mirror neuron inspired hierarchical network for action selection," *Proc. NeuroBotics*, pp. 89–97, 2004.

[8] S. Wermter, C. Weber, M. Elshaw, V. Gallese, and F. Pulvermuller, "A mirror neuron inspired hierarchical network for action selection," *Biomimetic neural learning for intelligent robots*, pp. 162–181, 2005.

[9] G. Metta, G. Sandini, L. Natale, L. Craighero, and L. Fadiga, "Understanding mirror neurons: a bio-robotic approach," *Interaction studies*, vol. 7, no. 2, pp. 197–232, 2006.

[10] A. Murata, L. Fadiga, L. Fogassi, V. Gallese, V. Raos, and G. Rizzolatti, "Object representation in the ventral premotor cortex (area f5) of the monkey," *Journal of Neurophysiology*, vol. 78, no. 4, p. 2226, 1997.

[11] M. Iacoboni, I. Molnar-Szakacs, V. Gallese, G. Buccino, J. C. Mazziotta, and G. Rizzolatti, "Grasping the intentions of others with one's own mirror neuron system," *PLoS Biol*, vol. 3, no. 3, p. e79, 02 2005.

[12] J. Grezes, J. L. Armony, J. Rowe, and R. E. Passingham, "Activations related to and neurones in the human brain: an fmri study," *NeuroImage*, vol. 18, no. 4, pp. 928 – 937, 2003.

[13] E. Oztop and M. Arbib, "Schema design and implementation of the grasp-related mirror neuron system," *Biological cybernetics*, vol. 87, no. 2, pp. 116–140, 2002.

[14] J. Bonaiuto, E. Rosta, and M. Arbib, "Extending the mirror neuron system model, i," *Biological cybernetics*, vol. 96, no. 1, pp. 9–38, 2007.

[15] J. Bonaiuto and M. Arbib, "Extending the mirror neuron system model, ii: what did i just do? a new role for mirror neurons," *Biological cybernetics*, vol. 102, no. 4, pp. 341–359, 2010.

[16] R. Memisevic and G. Hinton, "Unsupervised learning of image transformations," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[17] R. Memisevic and G. E. Hinton, "Learning to represent spatial transformations with factored higher-order boltzmann machines," *Neural Comput.*, vol. 22, pp. 1473–1492.

[18] G. Taylor, R. Fergus, Y. LeCun, and C. Bregler, "Convolutional learning of spatio-temporal features," *Computer Vision–ECCV 2010*, pp. 140–153, 2010.

[19] G. Taylor, G. Hinton, and S. Roweis, "Modeling human motion using binary latent variables," *Advances in neural information processing systems*, vol. 19, p. 1345, 2007.

[20] I. Sutskever and G. Hinton, "Learning multilevel distributed representations for high-dimensional sequences," in *Proceeding of the Eleventh International Conference on Artificial Intelligence and Statistics*. Citeseer, 2007, pp. 544–551.

[21] I. Sutskever, G. Hinton, and G. Taylor, "The recurrent temporal restricted boltzmann machine," *Advances in Neural Information Processing Systems*, vol. 21, 2009.

[22] G. Hinton, A. Krizhevsky, and S. Wang, "Transforming auto-encoders," in *Artificial Neural Networks and Machine Learning - ICANN 2011*, ser. Lecture Notes in Computer Science, T. Honkela, W. Duch, M. Girolami, and S. Kaski, Eds. Springer Berlin / Heidelberg, 2011, vol. 6791, pp. 44–51.

[23] J. Zhong, C. Weber, and S. Wermter, "Robot trajectory prediction and recognition based on a computational mirror neurons model," in *Artificial Neural Networks and Machine Learning - ICANN 2011*, ser. Lecture Notes in Computer Science, T. Honkela, W. Duch, M. Girolami, and S. Kaski, Eds. Springer Berlin / Heidelberg, 2011, vol. 6792, pp. 333–340.

[24] G. Hinton, "A practical guide to training restricted boltzmann machines," *Momentum*, vol. 9, p. 1, 2010.

[25] N. Aldebaran, "Robotics http://www. aldebaran-robotics. com/eng," *NaoRobocup. php [Consulta Julio 2009]*.

[26] L. Kneip, F. Tâche, G. Caprari, and R. Siegwart, "Characterization of the compact hokuyo urg-04lx 2d laser range scanner," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, 2009, pp. 1447–1454.