A Perceptual Memory System for Affordance Learning in Humanoid Robots

Marc Kammer[1,2], Marko Tscherepanow[1,2], Thomas Schack[1,3], and Yukie Nagai[1,4]

[1]CITEC, Cognitive Interaction Technology, Center of Excellence
[2]Applied Informatics, Faculty of Technology
[3]Neurocognition and Action, Faculty of Psychology and Sport Sciences
Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany

[4]Graduate School of Engineering, Osaka University,
2-1 Yamadaoka, Suita Osaka, 565-0871 Japan

mkammer@cit-ec.uni-bielefeld.de

Abstract. Memory constitutes an essential cognitive capability of humans and animals. It allows them to act in very complex, non-stationary environments. In this paper, we propose a perceptual memory system, which is intended to be applied on a humanoid robot learning affordances. According to the properties of biological memory systems, it has been designed in such a way as to enable life-long learning without catastrophic forgetting. Based on clustering sensory information, a symbolic representation is derived automatically. In contrast to alternative approaches, our memory system does not rely on pre-trained models and works completely unsupervised.

Keywords: cognitive robotics, artificial memory, life-long learning, affordances

1 Introduction

How humanoid robots can be enabled to learn complex real-world tasks is still an open research question. Recently, the concept of affordances has become a popular paradigm in teaching robots. The psychologist Gibson [6] defined the term affordances as action opportunities an observer becomes aware of by looking at an environment or at an object; for example, a car affords to drive and a ball affords to kick. The learning of affordances requires a robot to memorize certain types of objects, actions, effects as well as their relationships [5].

Since real-world environments are usually dynamic, an agent acting in them needs to adapt continuously. This requires the ability of life-long learning and the stable memorization of relevant information. Although important, there are only few investigations (e.g., [2]) on memory systems in cognitive robots that enable robots to learn, inter alia, affordances.

T. Honkela et al. (Eds.): ICANN 2011, Part II, LNCS 6792, pp. 349–356, 2011. The original publication is available at www.springerlink.com.

In this paper, we present the basis of a biologically inspired and distributed perceptual memory system for cognitive robots. It enables the stable, unsupervised, and incremental learning of perceptual information as well as the automatic generation of symbolic object representations. A symbolic and therefore discretized representation of perceptual information is a necessary requirement for learning affordances [13].

The rest of the paper is organized as follows: First, we will give an overview of biological and technical background information that forms the theoretical foundation of the introduced memory approach. Second, we present a prototypical implementation of the developed memory and will show first evaluation results using a real-world data set. Finally, we summarize the results and discuss challenges, possible improvements and potential future application scenarios.

2 Memory as a Basis for Learning in Cognitive Robotics

We claim that memory is a necessary requirement for any form of learning; as pointed out by Baxter and Browne in [2](p. 1), "cognition is inherently memory-based". In biological organisms even the most basic acquisition of knowledge is already a form of learning and it is difficult, if possible at all, to define where the process of memorization ends and the process of learning starts [10].

Neuroscientific and biological research offers a rich foundation of theories and models which can, if not in detail but in principle, be used to simulate cognitive capabilities such as memory by technical means. For instance, human memory can be divided into several subsystems regarding neural correlates, temporal aspects, and content [12].

In order to meet the capacity and time constraints of life-long learning systems, stored information needs to be organized efficiently, which is summarized by the term *cognitive economy* [7]. In particular, the amount of data must be reduced by mechanisms such as categorization to meet the storage and time requirements. In [4] the principle of cognitive economy is satisfied by using Self-Organizing Feature Maps to create a heteroassociative memory which learns categories based on prototype representations. But the network structure is not incremental and therefore limited in its capability to incorporate novel data.

A further problem arising in life-long learning systems is the *stability-plasticity dilemma* [8]: How can a system retain old memories but still learn new information? The memory system introduced in [11] approaches the stability-plasticity dilemma by adopting a biologically inspired hierarchical visual pathway processing. However the solution does not allow explicit symbolic access to the learned entities, which would be beneficial for applying reasoning methods in the context of learning affordances.

In [15], Sun discusses the technical representation of memorized information. He favors a combination of subsymbolic and symbolic representations, as realized within the cognitive architecture CLARION [16]. The CLARION architecture simulates human mental processes, which works in simulation but not in a real world environment.

Hanheide and his colleagues [9] presented a perceptual memory system for learning faces of interaction partners in a real-world scenario. It uses pre-trained models for face detection and feature extraction. Known interaction partners are classified by means of support vector machines following an one-vs-rest approach. If unknown persons appear, the corresponding face patches are used to train a new classifier. The one-vs-rest approach requires the extracted face patches of all known interaction partners to be stored, which is not compatible with cognitive economy and puts the scalability of the system in question.

Although related works from the field of affordance learning do not focus on memory aspects in an overall cognitive architecture as investigated in [2], they incorporate at least a very simple form of memory to represent objects from the real world. For example [13] and [18] use X-Means clustering for a perceptual discretization and memorization to apply a high level reasoning, which limits the number of objects that can be used severely.

As the goal of our architecture is the incremental and online learning of affordances using only few training samples the architectural memory has to fulfill the above mentioned criteria. Therefore, we introduce a perceptual memory system based on Adaptive Resonance Theory (ART) [8] networks, which learn online, incremental, unsupervised and constitute a solution of the *stability-plasticity dilemma*. Similar to the CLARION architecture, we use a subsymbolic feature processing mechanism to form a symbolic feature representation at the top level of the memory. A detailed explanation of the developed architecture is given in the next section.

3 Our Perceptual Memory System

We met the above-mentioned requirements by creating a distributed, hierarchical, incremental perceptual network structure, composed of different ART networks. These ART networks are capable of unsupervised incremental online learning and can be trained on few training samples. As mentioned before, the long term goal of this work is the creation of an interactive learning architecture that shall be used in an online affordances learning scenario. Each displayed constituent is described in this section.

Our so far preliminary cognitive architecture is able to detect objects that are introduced into a static scene by observing changes. Detected objects are rotated according to their first principal axis to mimic the process of *mental rotation* [14] and scaled to a common size. These normalized images are further processed according to Fig. 1. The normalized object patches are split into five partitions, as indicated by the vertical lines in Fig. 1. For each of these partitions several histograms reflecting the frequency of the occurrence of specific colors are determined. The colors are defined by ranges in the hue (H) plane of the HSV color space. Each range was defined manually and corresponds to the color impression it triggers in the human perceptual system. Finally, the histograms of all partitions that correspond to an individual color are concatenated and fed as input into an ART network. These networks learn efficient sub-symbolic

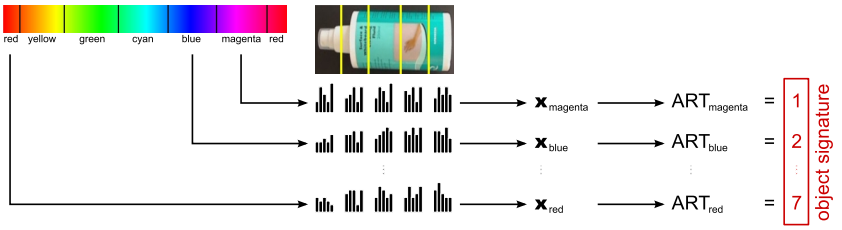

Fig. 1. Processing of the normalized object patches. After splitting an image into five disjunct partitions, histograms are computed for several color ranges. These color-specific histograms are concatenated and fed into ART networks. Finally, the indices of the best-matching node of each ART network are concatenated to yield object-specific symbolic signatures.

representations of presented input samples which are referred to as categories. For each presented object patch, the indices of the best-matching categories form an object-specific signature, which we regard as symbolic representation of an object. The common activation of all ART networks therefore provides a distributed, unique representation of an object image i, which we term *object signature*, referring to it as $\boldsymbol{s_i}$.

$$\boldsymbol{s_i} = (s_{i1}, s_{i2}, ..., s_{in}) \tag{1}$$

Such a distributed feature representation allows the system to identify certain features in reasoning processes as important or irrelevant. This is important, for example in the task of affordance learning. In order to measure the dissimilarity between two object signatures $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, we apply the hamming distance $\Delta(\boldsymbol{s_i}, \boldsymbol{s_j})$.

$$\Delta(\boldsymbol{s_i}, \boldsymbol{s_j}) = \frac{1}{n} \sum_{k=1}^{n} d(s_{ik}, s_{jk}) \quad \text{, with } \; d(s_{ik}, s_{jk}) = \begin{cases} 1 & \text{if } s_{ik} \neq s_{jk} \\ 0 & \text{if } s_{ik} = s_{jk} \end{cases} \tag{2}$$

The hamming distance $\Delta(\boldsymbol{s_i}, \boldsymbol{s_j})$ counts the number of positions at which two signatures differ. The results are normalized to the interval $[0, 1]$, where 0 means that both object signatures are identical and 1 means that both object signatures differ at all positions.

4 Experimental Results

We compared three different ART networks: Fuzzy ART [3], TopoART [17] and Hypersphere ART [1]. All of these networks allow for stable and incremental learning of new objects. But they differ in their activation functions and their sensitivity to noise. The goal of the evaluation process of our perceptual memory architecture is to identify the best parameters for each used ART-network, as

well as to evaluate the performance of each ART-network in the given scenario and if it is suited or not.

We used the described preliminary cognitive architecture to create a real world data set by recording 25 different objects with 10 variants for each object, which resulted in an overall set of 250 images. This set was split in disjunct training and test sets, each containing 125 images. Using a real world setup resulted in different variants for each object. In each recording attempt, the position of the object might vary as well as the lighting conditions or even noise can lead to different object appearances. Figure 2 shows five of the 25 different objects with all of its variants.

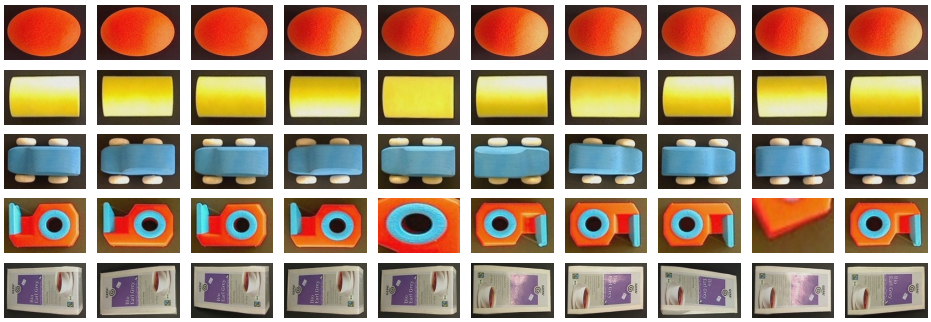

Fig. 2. Recorded and normalized images of five exemplary objects. For each object (rows) ten different images (columns) were acquired and processed automatically. In some very rare cases, the images were incorrectly cropped.

The best parameter values for ρ (all networks) and β_{sbm} (only TopoART) were determined by grid search. Due to the small amount of available data, all networks were trained in fast-learning mode (β=1) and the noise reduction mechanism of TopoART was disabled (φ=1,$\tau \geq$1). The selected parameters were used in the test phase.

As evaluation criteria, two different measures denoted by d_w and d_b were used. The first measure d_w, given in (4), represents the dissimilarity each object has within its variants. A low value is desirable, as it indicates that the signature of the variants of the objects are similar to each other. θ and ϑ denote the number of all objects and their variants, respectively. In our experiments, $\theta = 25$ and $\vartheta = 5$ were used for training as well as testing. Then the number of all hamming distances an object can have within its own variants is $\varrho = \frac{1}{2}\vartheta \cdot (\vartheta - 1)$. Then, the mean over all ϱ hamming distances for one object o is given by d_w^o.

$$d_w^o = \frac{1}{\varrho} \cdot \sum_{i=1}^{\vartheta} \sum_{j=i+1}^{\vartheta} \Delta\left(\boldsymbol{s_i^o}, \boldsymbol{s_j^o}\right) \tag{3}$$

If an specific object o has three variants ϑ=3, for example, an overall of $\varrho = 3$ hamming distances can be calculated which results to $d_w^o = \frac{1}{\varrho}\left(\Delta\left(\boldsymbol{s_1^o}, \boldsymbol{s_2^o}\right) +\right.$

$\Delta(\boldsymbol{s_1^o}, \boldsymbol{s_3^o}) + \Delta(\boldsymbol{s_2^o}, \boldsymbol{s_3^o})\big)$. The mean of d_w^o over all objects is denoted by d_w.

$$d_w = \frac{1}{\theta} \sum_{o=1}^{\theta} d_w^o \tag{4}$$

The second measure d_b, given in (6), represents the dissimilarity each object variant has to all variants of the other objects and is therefore termed the between object similarity. A high value indicates a high dissimilarity of variants between different objects which is preferable. The mean dissimilarity of a specific variant i of an object o to all variants of other objects is denoted by $d_b(\boldsymbol{s_i^o})$.

$$d_b(\boldsymbol{s_i^o}) = \frac{1}{(\theta - 1) \cdot \vartheta} \sum_{\substack{o'=1 \\ o' \neq o}}^{\theta} \sum_{j=1}^{\vartheta} \Delta(\boldsymbol{s_i^o}, \boldsymbol{s_j^{o'}}) \tag{5}$$

d_b denotes the mean dissimilarity averaged over all objects and their variants.

$$d_b = \frac{1}{\theta \cdot \vartheta} \sum_{o=1}^{\theta} \sum_{i=1}^{\vartheta} d_b(\boldsymbol{s_i^o}) \tag{6}$$

For parameter optimization, d_w should be low, while for d_b high values are desired. This relationship is captured by the goal function G.

$$G = d_b - d_w \tag{7}$$

For the maximum of G, each individual object is represented by a set of similar variant signatures, while the signatures of different objects differ strongly. The chosen parameters at the maximum of G are therefore optimal.

Figure 3 depicts an exemplary evaluation of the FuzzyART network, showing d_w, d_b, and the goal function G averaged over ten training trials depending on the vigilance parameter ρ. In addition, it compares the goal functions for all ART networks.

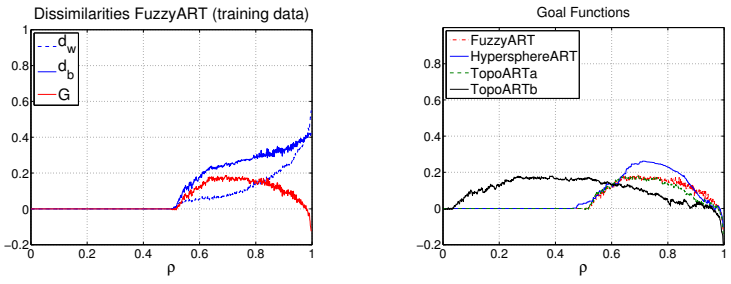

Fig. 3. Training results of an exemplary FuzzyART network (left) and the corresponding goal functions G of all ART networks (right) depending on ρ.

Table 1. Test results and their standard deviations calculated with the optimal parameter settings for each network. The settings for TopoART were independently optimized for its components TopoART a (I) and TopoART b (II).

	FuzzyART $\rho = 0.696$	HypersphereART $\rho = 0.71$	TopoART a/b (I) $\rho = 0.694$ & $\beta_{sbm} = 0$	TopoART a/b (II) $\rho = 0.388$ & $\beta_{sbm} = 0$
d_w	0.0598857	0.0221714	0.0626286 / 0.096	0 / 0.0626286
d_b	0.135429	0.117171	0.129229 / 0.240762	0 / 0.129229
σ_w	0.305495	0.216834	0.318894 / 0.393548	0 / 0.318894
σ_b	0.0097925	0.0155369	0.010532 / 0.0145529	0 / 0.010532

Finally, the estimated parameters were used in the test phase to calculate the hamming distances on the separated test set. The results summarized in Table 1 indicate that HypersphereART is best suited for the given setup as its difference of $d_b - d_w$ has the overall highest value which mean that variants of the same object are represented by similar signatures and that the signatures between different objects have a greater deviation.

5 Conclusion and Outlook

In this paper, we presented a distributed, incremental perceptual memory system tailored to the task of learning affordances in an interactive real-world scenario. In order to fulfill the requirement of stable life-long learning, we applied different ART networks (Fuzzy ART, Hypersphere ART, and TopoART) to learn subsymbolic object representations. Furthermore, we introduced the idea of using a distributed but common activation of different ART networks to create a bottom-up symbolic object representation based on the respective best-matching nodes.

In our experimental setup, Hypersphere ART performed best. Furthermore, the results of Fuzzy ART and TopoART are very similar to each other. Therefore, we conclude that the Euclidean distance used for activating nodes in Hypersphere ART is more suited to the task at hand than the city block distance utilized by Fuzzy ART and TopoART.

Our future research will focus on additional perceptual input features for categorizing objects and enriching the object signature, e.g. by shape and size information. Also especially a recognition test, based on an enlarged object signature and a distance measurement, as for example the used hamming distance has to be investigated. Another future investigation is the comparison of the ART networks with networks that suffice most of the required criteria but do not belong to the ART family, as for example growing self organizing maps or growing neural gas.

Acknowledgements. This research was funded by the German Research Foundation (DFG), Excellence Cluster 277, Cognitive Interaction Technology, Research Area D (Memory and Learning).

References

1. Anagnostopoulos, G.C., Georgiopoulos, M.: Hypersphere ART and ARTMAP for unsupervised and supervised incremental learning. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN). vol. 6, pp. 59–64 (2000)
2. Baxter, P., Browne, W.: Memory as the substrate of cognition: A developmental cognitive robotics perspective. In: Johansson, B., Sahin, E., Balkenius, C. (eds.) Proceedings of the International Conference on Epigenetic Robotics (EpiRob). pp. 19–26 (2010)
3. Carpenter, G.A., Grossberg, S., Rosen, D.B.: Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. Neural Networks 4(6), 759–771 (1991)
4. Chartier, S., Giguère, G., Langlois, D.: A new bidirectional heteroassociative memory encompassing correlational, competitive and topological properties. Neural Networks 22(5–6), 568–578 (2009)
5. Şahin, E., Çakmak, M., Doğar, M.R., Uğur, E., Üçoluk, G.: To afford or not to afford: A new formalization of affordances toward affordance-based robot control. Adaptive Behavior 15(4), 447–472 (2007)
6. Gibson, J.J.: The theory of affordances. Perceiving, Acting, and Knowing pp. 67–82 (1977)
7. Goldstone, R.L., Kersten, A.: Concepts and categorization. In: Healy, A.F., Proctor, R.W. (eds.) Comprehensive handbook of psychology, vol. 4: Experimental psychology, pp. 599–621. Wiley (2003)
8. Grossberg, S.: Competitive learning: From interactive activation to adaptive resonance. Cognitive Science 11, 23–63 (1987)
9. Hanheide, M., Wrede, S., Lang, C., Sagerer, G.: Who am I talking with? A face memory for social robots. In: IEEE International Conference on Robotics and Automation. pp. 3660–3665. IEEE (2008)
10. Kandel, E.R.: The molecular biology of memory storage: a dialogue between genes and synapses. Science 294(5544), 1030–1038 (2001)
11. Kirstein, S., Wersing, H., Körner, E.: A biologically motivated visual memory architecture for online learning of objects. Neural Networks 21(1), 65–77 (2008)
12. Markowitsch, H.J., Halligan, P.W., Kischka, U., Marshall, J.C.: Functional neuroanatomy of learning and memory, pp. 724–741. University Press (2003)
13. Montesano, L., Lopes, M., Bernardino, A., Santos-Victor, J.: Learning Object Affordances: From Sensory–Motor Coordination to Imitation. IEEE Transactions on Robotics 24(1), 15–26 (2008)
14. Shepard, R.N., Metzler, J.: Mental rotation of three-dimensional objects. Science pp. 701–703 (1971)
15. Sun, R.: Robust reasoning: Integrating rule-based and similarity-based reasoning. Artificial Intelligence 75(2), 241–295 (1995)
16. Sun, R., Zhang, X., Mathews, R.: Capturing human data in a letter counting task: Accessibility and action-centeredness in representing cognitive skills. Neural Networks 22(1), 15–29 (2009)
17. Tscherepanow, M.: TopoART: A topology learning hierarchical ART network. In: Proceedings of the International Conference on Artificial Neural Networks (ICANN). LNCS, vol. 6354, pp. 157–167. Springer (2010)
18. Ugur, E., Sahin, E., Oztop, E.: Affordance learning from range data for multi-step planning. In: Proceedings of the International Conference on Epigenetic Robotics (EpiRob). pp. 177–184 (2009)