Affordance-based object recognition using interactions obtained from a utility maximization principle

Tobias Kluth, David Nakath, Thomas Reineking, Christoph Zetzsche, and Kerstin Schill

Cognitive Neuroinformatics, University of Bremen, Enrique-Schmidt-Straße 5, 28359 Bremen, Germany,
tkluth@math.uni-bremen.de,
WWW home page: http://www.informatik.uni-bremen.de/cog_neuroinf/en

Abstract. The interaction of biological agents within the real world is based on their abilities and the affordances of the environment. By contrast, the classical view of perception considers only sensory features, as do most object recognition models. Only a few models make use of the information provided by the integration of sensory information as well as possible or executed actions. Neither the relations shaping such an integration nor the methods for using this integrated information in appropriate representations are yet entirely clear. We propose a probabilistic model integrating the two information sources in one system. The recognition process is equipped with an utility maximization principle to obtain optimal interactions with the environment. We compared an affordance-based system to a non-affordance-based one, both relying on the same architecture, in a simulated and a real world scenario.

Keywords: affordance, sensorimotor, object recognition, Bayesian inference, information gain

1 Introduction

The ability of humans to reliably recognize objects in the environment is still a largely unsolved problem for artificial systems. Usually, object recognition is understood as a classification problem where a fixed mapping from feature vectors to object classes is learned. This is in line with the classical view of perception as the input from world to mind and of action as the output from mind to world [1], which implies a dissociation between the capacities for perception and action. However, there is strong evidence that object recognition cannot be understood independently of the interaction of an agent with its environment [2]. “Active perception” approaches [3–5] take this partially into account, but actions are not merely means for acquiring new information, they also provide evidence themselves for the recognition [6]. What an agent perceives is thus also determined by what it does or what it is able to do [2].

Research in the direction of affordances by Gibson [7, 8], see [9] for a comprehensive overview, also gives evidence that affordances are key ingredients of the perceptual process. In Gibson's opinion an agent and its environment complement each other such that it is necessary to study the agent in its natural environment rather than in isolation. A variety of studies regarding object recognition show that the visual information of a manipulable object causes an activation of representations of actions which can typically be executed on the object [10–12], cf. Fig. 1. Moreover, Tucker & Ellis [12], for example, found that the affordance-based compatibility effect in a grasping task is statistically indistinguishable when an object is represented by a visual stimulus or by its written name. They concluded that on-line visual processing is not necessary to generate affordance-based compatibility effects. The link between object representations and actions appears to be critical as the associated actions can be activated independent of the mode of representation. This gives additional evidence that an object is not only encoded by the sensor features but also by the (possibly) applied manipulations [6]. The advantageous interplay between sensory and action information, which was also recognized by Neisser [13], should be considered in the recognition process. In cognitive models affordances were related to low-level processes [14] as well as they were considered to be part of a complete cognitive model [15, 16]. In robotics the theory of affordances is mostly used for behavior-based control of robots [17]. The learning of affordances was considered in mainly two different aspects. On the one hand, learning the consequences of an action in a given situation [18–21] and on the other hand learning the properties of the environment which afford a specific behavior [22–24].

The strong interrelation between motor actions and sensory perceptions is basis for the concept of a sensorimotor representation [2, 25–27]. Similarly to the affordance point of view the processing stages for sensory and motor information are not separated. Strictly speaking, this is a precondition for a sensorimotor representation which is obtained from alternating sensory perceptions and motor actions [25, 28]. The approach including the actions in the representation gives the opportunity to choose the next action such that a specific objective is pursued. Generally, the problem of action selection can be solved in numerous ways, but as information gathering should be one major purpose of motor actions it is interesting to consider an information-theoretic utility function. Prior research in this area often found that the principle of *information gain* is well suited to select an appropriate next action. This has been shown by [29] in the context of decision trees, where information gain was used to decide which attributes are the most relevant ones. The information gain strategy was able to model human behavior [30, 31] and could be used to mimic it in the case of human-like expert systems [32, 33] and saccadic eye movements [34, 35]. In robotics this utility function was also successfully applied to uncertainty minimization, for example in robot localization tasks [36, 37].

In this paper, we propose a system for object recognition which incorporates both the information gain principle from sensorimotor systems and the theoretical concept of affordances. Building upon the investigations in [34, 35], we

Fig. 1: A sequence of interactions with an object (coffee pot) in the real world is shown. The illustrated affordances of the object are "open it" (middle) and "hold it on the handle" (right).

developed a sensomotoric probabilistic reasoning system for affordance-based object recognition. The design of our architecture is motivated by two main goals: i) to provide a clear relation to Bayesian inference approaches, and ii) to enable a comparison between the classic sensory approach and the sensorimotor, affordance-oriented approach within one common probabilistic framework. For this, we have generalized the original interpretation of sensomotoric features in terms of feature-action-feature triples as suggested in [34, 35].

The proposed system for object recognition has two main characteristics. First, actions are an explicit part of the representation. Instead of just using the sensory information for representing an object class, the system learns the joint distribution over sensory features and corresponding actions, resulting in a sensorimotor representation of objects. This enables the system to perform actions that are most appropriate for recognizing a particular type of object. The second characteristic of the system is its ability to choose actions that are optimal for *discerning* different types of objects. This exceeds the typical affordance view because the system initially does not know what type of object it is interacting with. The uncertainty about the object class requires the consideration of all possible object types for the action selection. This selection is based on maximizing the expected information gain associated with an action, resulting in optimal behavior in terms of the quickest possible reduction of uncertainty.

The basic architecture of the system we propose is outlined in Sect. 2. In Sect. 3, we describe the implementation of the system. Sect. 4 shows the results of simulated and real world data sets and compares an affordance-based object recognition approach to a standard sensor-based approach. Furthermore the adaption-effect of the information gain strategy is quantified. The paper is concluded with a discussion of the advantages offered by the consideration of affordances within the system.

2 Object Recognition System

The system described in the following is a generic architecture (see Fig. 2). In the case of affordance-based object recognition, the recognition loop starts out with a particular pose of an object which is perceived by a sensor. It subsequently passes its raw sensor data to the sensory processing module. After processing, the sensory data becomes part of a new sensorimotor feature, which is then fed into the probabilistic reasoning module. The processed sensory data are also used to obtain a set of possible interactions, i.e., the affordances offered by the sensory data (related to the abilities of the agent). The Bayesian inference module calculates the new posterior distribution based on a previously-learned sensorimotor representation. This representation contains the learned perceptual consequences of an interaction in a given state for every object class. The posterior distribution constitutes the current belief of the system. This belief is used by the information gain strategy to choose an optimal next action from the set of possible interactions. The selected interaction then also becomes part of the sensorimotor feature and is subsequently executed by the agent. The whole process results in a new state, which in turn delivers new raw sensory data to enter the next cycle of the recognition loop.

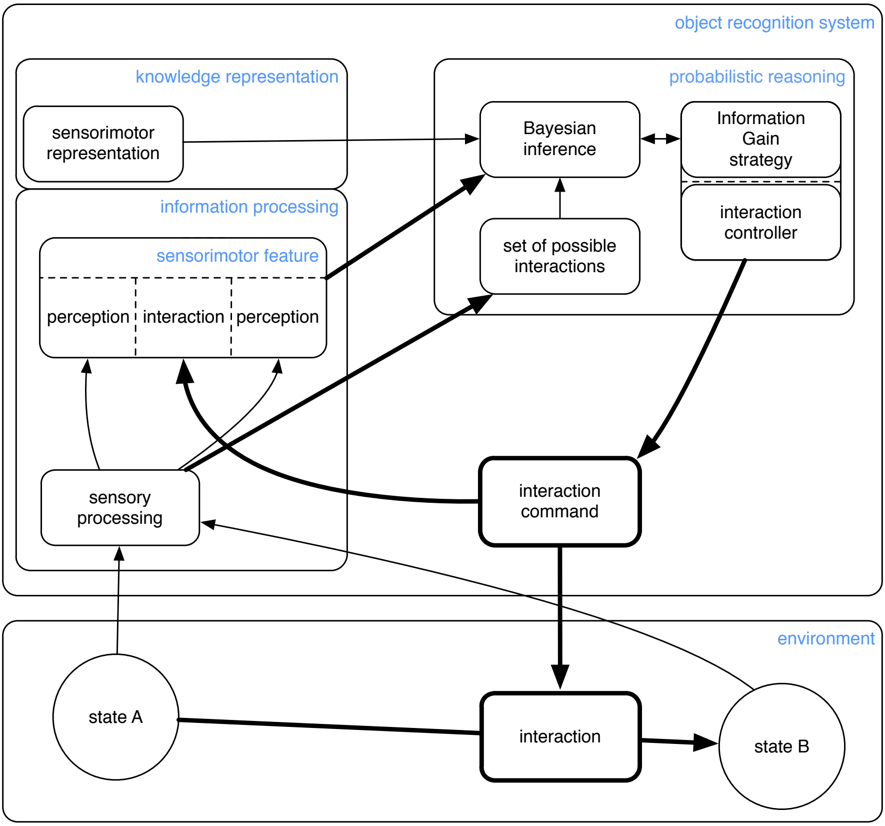


Fig. 2: Architecture of the object recognition system.

More formally spoken, the system depends on an *agent*, which can be controlled such that it perceives information about a specific aspect of the world. In Fig. 2, the two arrows pointing from the states to the sensory processing module correspond to the mapping $A : U \times X \to R$, where U is the space of all interactions which are currently possible, X is the state space, and R is the raw sensor data space.

The system has no explicit knowledge about the actual state, and the currently possible interactions U. The possible interactions are of course dependent on the state but nevertheless both information must be obtained from the sensor data. The sensoric dependency on the state is formalized by the mapping $U : X \to \mathcal{P}(\Omega_U)$, where Ω_U is the set of all possible interactions and $\mathcal{P}$ denotes the power set. Note that U comprises the link from the state to the sensory processing module and the following link to the set of possible interactions in Fig. 2, i.e., the perceived affordances. Assuming that the output of the function U is given, we write U instead of $U(x),\ x \in X$, for convenience. Considering the state-agnostic behavior, the influence of the agent can be formally redefined to $A_x : U \to R$ where the index x recalls the dependency on the state

$$A_x(u) := A(x, u) = r,\ x \in X,\ u \in U(x),\ r \in R. \tag{1}$$

The only time-dependent variables are the state x and the interaction u.

The raw sensor data $r \in R$ is fed into the *sensory processing* (SP) which mainly extracts the relevant features belonging to a feature space F, i.e., $SP : R \to F$. Subsequently, the quantization operation $Q_S : F \to S$ maps the features to a specific feature class in the finite and countable space S. The possible interactions are mapped with $Q_M : \Omega_U \to M$ to the finite countable set of interactions M to yield a manageable product space of sensory information and actions. The results of these quantizations then become part of a sensorimotor feature (SMF). The single quantizations are represented in Fig. 2 by the arrows from the sensory processing module and the interaction command to the first-order sensorimotor feature which is defined as the triple

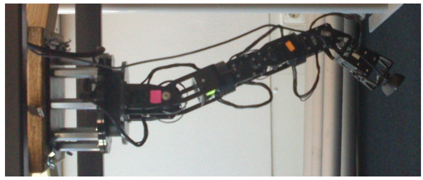
$$SMF_i := \{s_{i-1}, m_{i-1}, s_i\}, \tag{2}$$

where $m_{i-1} := Q_M(u_{i-1})$ is the interaction between the sensor information s_{i-1} and s_i at time step t_{i-1} and t_i (see Fig. 3). The whole chain of operations to obtain the sensor information at a time step t_i can be described by

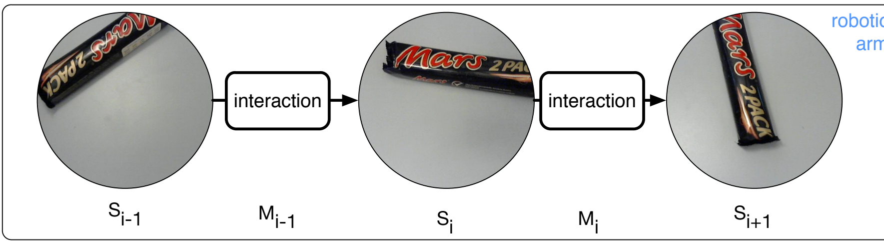
$$s_i := (Q_S \circ SP \circ A_x)(u_{i-1}). \tag{3}$$

The *knowledge representation* is comprised of the learned sensorimotor representation (SMR), which is a full joint probability distribution of $SMFs$ and the classes represented by the discrete random variable Y. Every possible SMF is generated on a set of known objects in a training phase. This means that, from every possible state x, the sensory consequence from every possible action u is perceived, resulting in

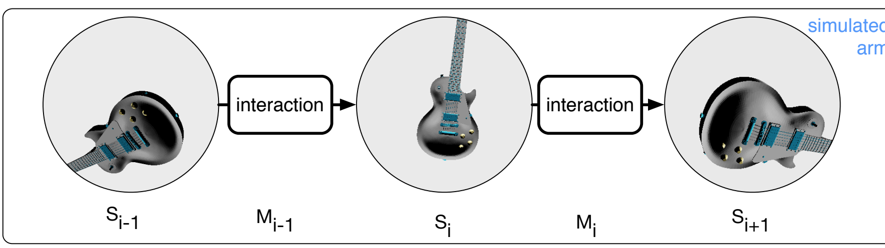
$$SMR := P(SMF, Y) = P(S_{i-1}, M_{i-1}, S_i, Y). \tag{4}$$



(a) Robotic arm



(b) Robotic arm object interactions



(c) Simulated object interactions

Fig. 3: The utilized robotic arm is shown in (a). Exemplary object interactions for the real case are shown in (b) and for the simulated case are shown in (c). Here, s_{i-1} denotes the preceding sensory input, m_{i-1} denotes the preceding action, and s_i denotes the current sensory input.

The *probabilistic reasoning* module consists of a Bayesian inference approach accompanied by an information gain strategy. They rely on bottom-up sensory data and top-down information from the knowledge representation. This design enables the Bayesian inference system to take into account interactions, thus improving the posterior distribution over the object classes Y. Furthermore, the information gain strategy can choose an optimal next interaction for the agent, thus improving the input of the following Bayesian inference step.

3 Model Implementation

Based on the schematic outline presented above, we applied our system in the field of object recognition. We consider both the case of a real (see Fig. 3b) and a simulated (see Fig. 3c) agent in form of a robotic arm interacting with objects in 3D space.

3.1 Agent Implementation

We used a discrete set of interactions M of a robotic arm with an object, both for the real arm and the simulated arm case. Hence, in both cases holds $\Omega_U = M$ and the quantization Q_M is an identity operation.

The following basic learning and recognition principles can be applied to both of the implemented agents: In the learning phase, features are extracted from the training data (i.e., images from every reachable state), which corresponds to the mapping SP introduced above. GIST-features [38, 39] are used to describe the sensory input which is comprised of holistic views on an object. The quantization Q_S is then learned by performing a k-means clustering on the extracted features ($k = 15$)[1]. In order to build the individual $SMFs$, features are extracted (by SP) and the results are assigned to clusters with the aid of the previously defined mapping Q_S. These labels are combined with the corresponding interactions in a set of $SMFs$. Finally, all generated $SMFs$ are stored in a Laplace-smoothed SMR.

3.2 Probabilistic Reasoning

The probabilistic reasoning is comprised of a Bayesian inference module in the form of a dynamic Bayesian network (BN) and a corresponding information gain strategy. Two of these probabilistic reasoning modules were implemented to examine the difference between *sensor networks*, which only take into account sensory information (which also implies that no information gain strategy is used), and *affordance-based networks*, which integrate sensory perceptions and interactions. The object recognition in the sense of computer vision then takes place by classification which is performed by choosing the class with the maximum posterior probability.

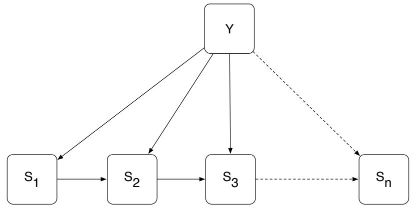
The representative of the *sensor networks* is Bayesian network 1 (BN1) (see Fig. 4a), which resembles an extended naive Bayes model by taking into account the current sensory input s_i and additionally assuming statistical dependencies between the preceding and the current sensor information, s_{i-1} and s_i, resulting in

$$P(y|s_{1:n}) = \alpha P(y)P(s_1|y)\prod_{i=2}^{n} P(s_i|s_{i-1}, y), \tag{5}$$

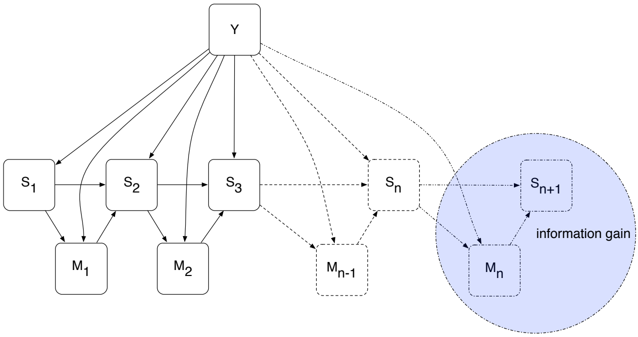
where α is a normalizing constant guaranteeing that the probability function properties are satisfied and $s_{1:n}$ is a short notation for the n-tuple $(s_1, \ldots, s_n)$.

Bayesian network 2 (BN2) (see Fig. 4b) uses the full information of the SMF and therefore belongs to the *affordance-based networks*. The assumption that the

[1] We use only a small number of clusters in order to limit the number of model parameters and to prevent overfitting.



(a) BN1



(b) BN2

Fig. 4: The dynamic Bayesian sensor network BN1, which is shown in (a), processes only sensory information. It represents a naive Bayes approach additionally assuming statistical dependencies between the preceding sensory input S_{i-1} and the current sensory input S_i. BN2, which is shown in (b), is an affordance-based dynamic Bayesian network. It assumes that the current sensory input S_i statistically depends on the object hypothesis Y, the preceding action M_{i-1}, and the preceding sensory input S_{i-1}. Additionally, it is assumed that the preceding action M_{i-1} statistically depends on the preceding sensory input S_{i-1} as well as on the object hypothesis Y.

current sensory input s_i depends on the preceding sensory input s_{i-1} and the intermediary action m_{i-1} integrates and sensory perceptions and actions in the recognition process and permits the application of the information gain strategy to choose the next optimal interaction. Additionally, it is assumed that the action m_{i-1} statistically depends on the preceding sensory input s_{i-1} and the object hypothesis Y, thus integrating the learned object affordances. The inference can then be conducted by

$$P(y|s_{1:n}, m_{1:n-1}) = \alpha P(y) P(s_1|y) \prod_{i=2}^{n} P(s_i|s_{i-1}, m_{i-1}, y) P(m_{i-1}|s_{i-1}, y). \quad (6)$$

3.3 Information Gain

The strategy for action selection should satisfy two main properties: (i) The strategy should adapt itself to the current belief state of the system and (ii) the strategy should not make decisions in an heuristic fashion but tightly integrated into the axiomatic framework used for reasoning. The information gain strategy presented in this paper complies with both of these properties.

The information gain IG of a possible next action m_n is defined as the difference between the current entropy $H(Y)$ and the conditional entropy $H(Y|S_{n+1}, m_n)$, i.e.,

$$IG(m_n) := H(Y) - H(Y|S_{n+1}, m_n), \tag{7}$$

where all probabilities are conditioned by $s_{1:n}, m_{1:n-1}$. This is equivalent to the mutual information of Y and (S_{n+1}, m_n) for an arbitrary m_n. As the current entropy $H(Y)$ is independent of the next action m_n the most promising action m^* can be calculated by minimizing the expected entropy with respect to S_{n+1}, i.e.,

$$m^* = \arg\min_{m_n}(\underset{S_{n+1}}{E}[H(Y|s_{1:n}, S_{n+1}, m_{1:n})]). \tag{8}$$

Because the sensory input s_{n+1} is not known prior to executing m_n, the expected value over all possible sensory inputs S_{n+1} is taken into account. Subsequently, the so chosen action $m^* \in M$ can be integrated into the sensorimotor feature. The inverse mapping of Q_M can then be used to obtain a top-down interaction command $u \in U$, which is then executed by the agent.

4 Evaluation

The implementation was evaluated on two datasets based on a k-fold cross validation scheme with $k = 10$. The case of the robotic arm movements can be seen as a realistic test for robustness with noisy ineractions and sensor data. This setting yielded a dataset, consisting of 8 object classes, each containing 10 objects from 30 different points of view.

The same setting was simulated resulting in a dataset consisting of 7 object classes, each containing 10 objects from 30 different points of view.

Figure 5 depicts the results of the robotic arm case. The integration of information gain-guided actions in the affordance-based network (BN2 +IG) proves to be beneficial in terms of recognition performance (see Fig. 5a). The sensor network BN1 performs worse, which holds true for the recognition performance as well as for the mean entropy reduction (see Fig. 5b). To particularly illustrate the effect of the information gain strategy, the affordance-based network performing information gain-guided actions (BN2 +IG) and random actions (BN2 -IG) were compared to each other. The sensorimotor networks with information gain-guided actions perform better (see Fig. 5a), which is reflected by a steeper reduction in entropy (see Fig. 5b).

In the simulated arm evaluation (see Fig. 6) the advantage of using affordance-based networks with information gain-guided actions (BN2 +IG) for recognition persists over most of the actions (see Fig. 6a). Only within the last five actions BN2 -IG is able to reach the same performance as the affordance-driven BN2 +IG. The affordance-agnostic BN1 is able to keep the pace with BN2 -IG within the first 15 actions but performs worse from this point on. The results in the performance domain are confirmed by the corresponding evolution of the mean entropies plotted in Fig. 6b.

The evaluation showed a comparison between affordance-based and affordance-agnostic object recognition approaches within the same architecture. Furthermore, the optional application of the information gain-strategy showed the effect of adapting an agent to the given affordances of an object. We found that integrating the information embedded in the affordances of an object is beneficial for the object recognition process. When our system additionally adapts itself to the given affordances the recognition process is significantly speeded up and sightly improved.

5 Conclusion

We have developed an affordance-based sensorimotor object recognition system. The architecture of this system tightly integrates action, perception, and reasoning and can thus make use of the information to be found in objects' affordances as well as adapts itself to them. The interaction with an object is driven by the principle of maximum information gain, where the system selects successively among the actions possible in each configurational state with the goal of minimizing the uncertainty about the object it is confronted with. A basic prerequisite for the proposed architecture is the description of objects with features which integrate sensory information as well as actions.

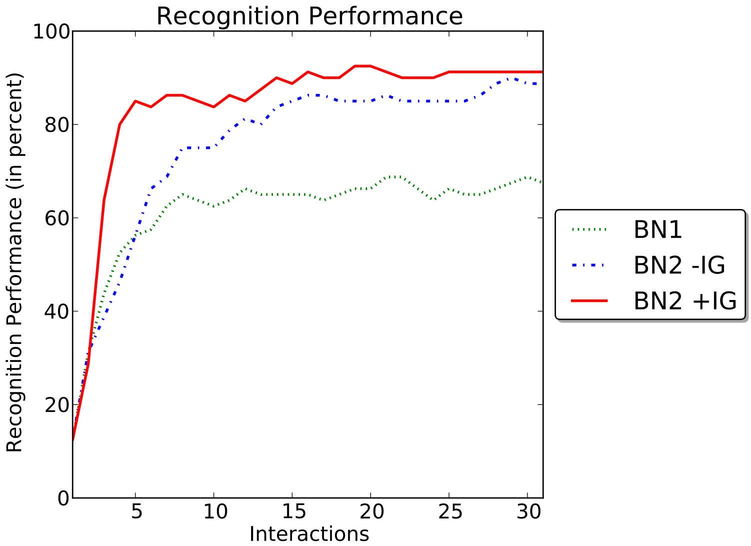
The evaluation of our system showed that the integration of affordance-information improved the recognition performance. When the system additionally adapts itself to the given object by choosing appropriate actions the performance is further increased. It could be shown that the proposed information gain strategy is well suited to control such an adaption process.

Currently the affordances are defined by a schematic learning process but our system design is not confined to such an approach. In the future, an improved learning process based on a detector for salient affordances will be implemented. Additionally, our system design also allows for different sensory modalities as it can process arbitrary sensory inputs. This complies with O'Regan who has a notion of sensorimotor contingencies which are not confined to particular sensory modalities [25, 28]. Thus, in future research tactile information should also be considered as sensory input.

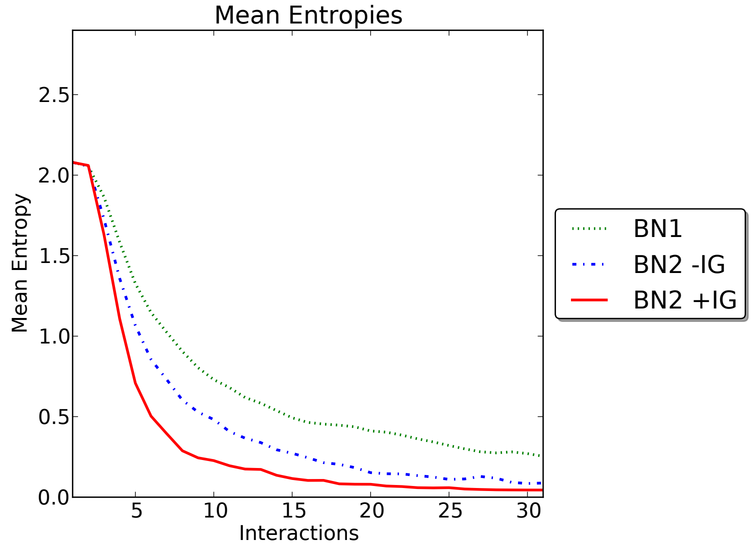
Acknowledgements This work was supported by DFG, SFB/TR8 Spatial Cognition, Project A5-[ActionSpace].

References

1. Hurley, S.L.: Consciousness in action. Harvard University Press (2002)

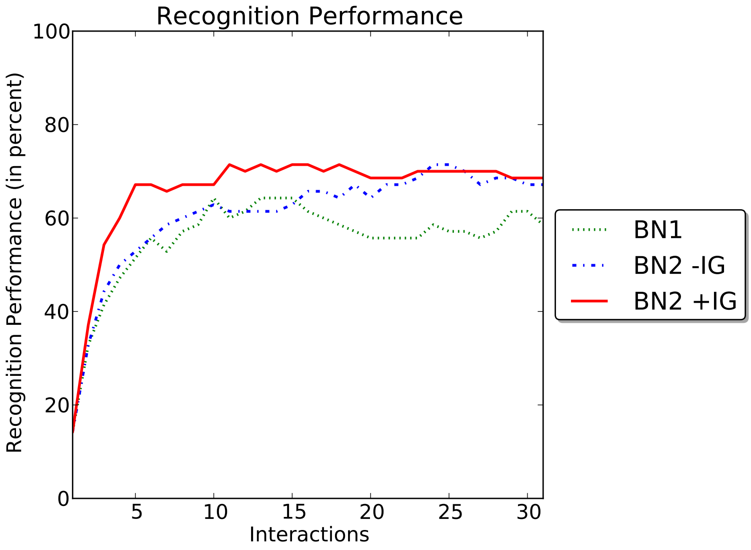


(a) Recognition performance

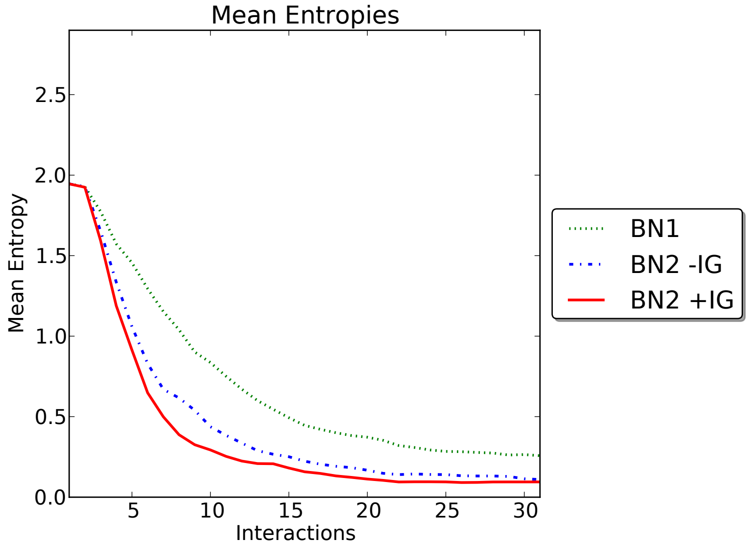


(b) Mean entropies

Fig. 5: Results of the robotic arm evaluation. BN 1 and 2 -IG executed random movements while BN2 +IG executed information gain-guided movements (GIST features, 15 clusters, 94 possible relative actions, inhibition of return). Recognition performance shown in (a) and mean entropy of the posterior distribution shown in (b) are both plotted against the number of performed interactions.



(a) Recognition performance



(b) Mean entropies

Fig. 6: Results of the simulated arm evaluation. BN 1 and 2 -IG executed random movements, while BN2 +IG executed information gain-guided movements (GIST features, 15 clusters, 94 possible relative actions, inhibition of return). Recognition performance shown in (a) and mean entropy of the posterior distribution shown in (b) are both plotted against the number of performed interactions.

2. Noë, A.: Action in Perception. MIT Press (2004)
3. Bajcsy, R.: Active perception. Proceedings of the IEEE **76**(8) (1988) 966–1005
4. Aloimonos, J., Weiss, I., Bandyopadhyay, A.: Active vision. International Journal of Computer Vision **1**(4) (1988) 333–356
5. Ballard, D.H.: Animate vision. Artificial intelligence **48**(1) (1991) 57–86
6. Helbig, H.B., Graf, M., Kiefer, M.: The role of action representations in visual object recognition. Experimental Brain Research **174**(2) (2006) 221–228
7. Gibson, J.J.: The theory of affordances. In R, S., J, B., eds.: Perceiving, acting, and knowing: toward an ecological psychology. Erlbaum, Hilsdale, NJ (1977) 67–82
8. Gibson, J.: The ecological approach to visual perception. Boston: Houghton Mifflin (1992)
9. Greeno, J.G.: Gibson's affordances. Psychological Review **101**(2) (1994) 336–342
10. Grèzes, J., Decety, J.: Does visual perception of object afford action? evidence from a neuroimaging study. Neuropsychologia **40**(2) (2002) 212–222
11. Pavese, A., Buxbaum, L.J.: Action matters: The role of action plans and object affordances in selection for action. Visual cognition **9**(4-5) (2002) 559–590
12. Tucker, M., Ellis, R.: Action priming by briefly presented objects. Acta psychologica **116**(2) (2004) 185–203
13. Neisser, U.: Cognition and reality: Principles and implications of cognitive psychology. WH Freeman/Times Books/Henry Holt & Co (1976)
14. Norman, J.: Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. Behavioral and brain sciences **25**(01) (2002) 73–96
15. Neisser, U.: Multiple systems: A new approach to cognitive theory. European Journal of Cognitive Psychology **6**(3) (1994) 225–241
16. Gibson, E.J.: Perceptual learning in development: Some basic concepts. Ecological Psychology **12**(4) (2000) 295–302
17. 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) (2007) 447–472
18. Fitzpatrick, P., Metta, G., Natale, L., Rao, S., Sandini, G.: Learning about objects through action-initial steps towards artificial cognition. In: Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on. Volume 3., IEEE (2003) 3140–3145
19. Stoytchev, A.: Behavior-grounded representation of tool affordances. In: Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on, IEEE (2005) 3060–3065
20. Stoytchev, A.: Toward learning the binding affordances of objects: A behavior-grounded approach. In: Proceedings of AAAI Symposium on Developmental Robotics. (2005) 17–22
21. Kjellström, H., Romero, J., Kragić, D.: Visual object-action recognition: Inferring object affordances from human demonstration. Computer Vision and Image Understanding **115**(1) (2011) 81–90
22. MacDorman, K.F.: Responding to affordances: Learning and projecting a sensorimotor mapping. In: Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on. Volume 4., IEEE (2000) 3253–3259
23. Cos-Aguilera, I., Canamero, L., Hayes, G.M.: Motivation-driven learning of object affordances: First experiments using a simulated khepera robot. In: In Proceedings of the 9th International Conference in Cognitive Modelling (ICCM03). (2003) 57–62

24. Cos-Aguilera, I., Hayes, G., Cañamero, L.: Using a sofm to learn object affordances. In: Proceedings of the 5th Workshop of Physical Agents (WAF'04), Girona, Catalonia, Spain. (2004)
25. O'Regan, J.K., Noë, A.: A sensorimotor account of vision and visual consciousness. Behavioral and brain sciences **24**(5) (2001) 939–972
26. Prinz, W.: A common coding approach to perception and action. Springer (1990)
27. Hommel, B., Müsseler, J., Aschersleben, G., Prinz, W.: The theory of event coding (TEC): A framework for perception and action planning. Behavioral and brain sciences **24**(05) (2001) 849–878
28. O'Regan, J.K.: What it is like to see: A sensorimotor theory of perceptual experience. Synthese **129**(1) (2001) 79–103
29. Quinlan, J.R.: Induction of decision trees. Machine learning **1**(1) (1986) 81–106
30. Oaksford, M., Chater, N.: Information gain explains relevance which explains the selection task. Cognition **57**(1) (1995) 97–108
31. Friston, K., Kilner, J., Harrison, L.: A free energy principle for the brain. Journal of Physiology-Paris **100**(1) (2006) 70–87
32. Schill, K., Pöppel, E., Zetzsche, C.: Completing knowledge by competing hierarchies. In: Proceedings of the Seventh conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers Inc. (1991) 348–352
33. Schill, K.: Decision support systems with adaptive reasoning strategies. In: Foundations of Computer Science, Springer (1997) 417–427
34. Zetzsche, C., Schill, K., Deubel, H., Krieger, G., Umkehrer, E., Beinlich, S.: Investigation of a sensorimotor system for saccadic scene analysis: an integrated approach. In: Proceedings of the 5th International Conference of Simulation of Adaptive Behaviour. Volume 5. (1998) 120–126
35. Schill, K., Umkehrer, E., Beinlich, S., Krieger, G., Zetzsche, C.: Scene analysis with saccadic eye movements: top-down and bottom-up modeling. Journal of Electronic Imaging **10**(1) (2001) 152–160
36. Cassandra, A.R., Kaelbling, L.P., Kurien, J.A.: Acting under uncertainty: Discrete Bayesian models for mobile-robot navigation. In: Proceedings of the 1996 IEEE/RSJ International Conference on Intelligent Robots and Systems' 96, IROS 96. Volume 2., IEEE (1996) 963–972
37. Stachniss, C., Grisetti, G., Burgard, W.: Information Gain-based Exploration Using Rao-Blackwellized Particle Filters. In: Robotics: Science and Systems. Volume 2. (2005) 65–72
38. Oliva, A., Torralba, A.: Building the gist of a scene: The role of global image features in recognition. Progress in brain research **155** (2006) 23–36
39. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision **42**(3) (2001) 145–175