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

SPATIAL COGNITION
A5-[ActionSpace]

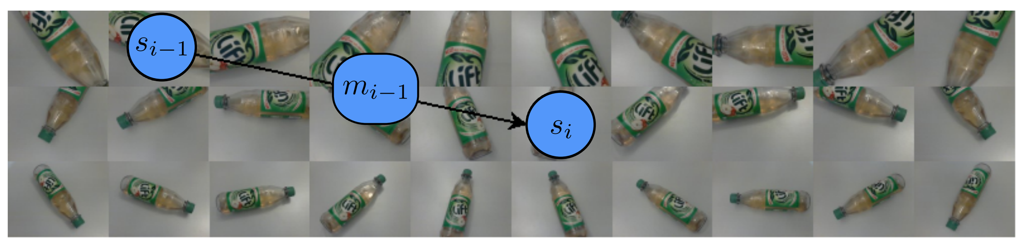
Tobias Kluth, David Nakath, Thomas Reineking, Christoph Zetzsche, Kerstin Schill

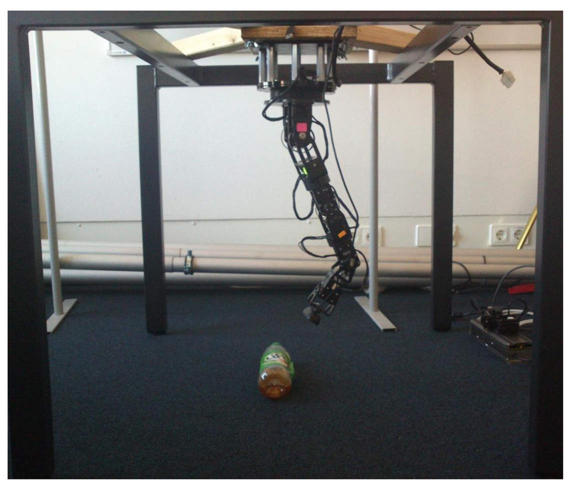
Introduction

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.

Sensorimotor Representation

Sensorimotor Feature:
$SMF_i := \{s_{i-1}, m_i, s_i\}$

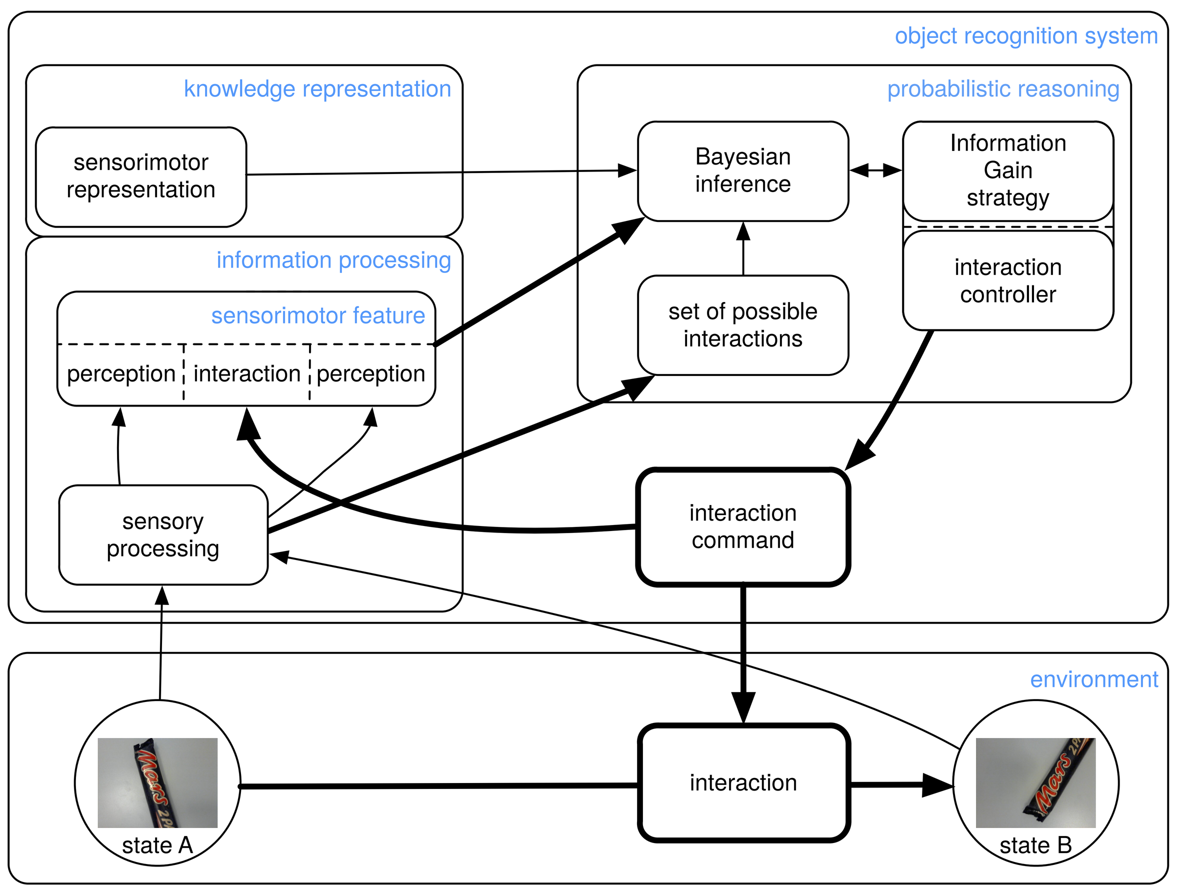




The *knowledge representation* is comprised of the learned sensorimotor representation (SMR), which is a full joint probability distribution of SMFs and the classes represented by the discrete random variable Y. Every possible SMF is generated on a set of known objects in a training phase. This means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

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

Model



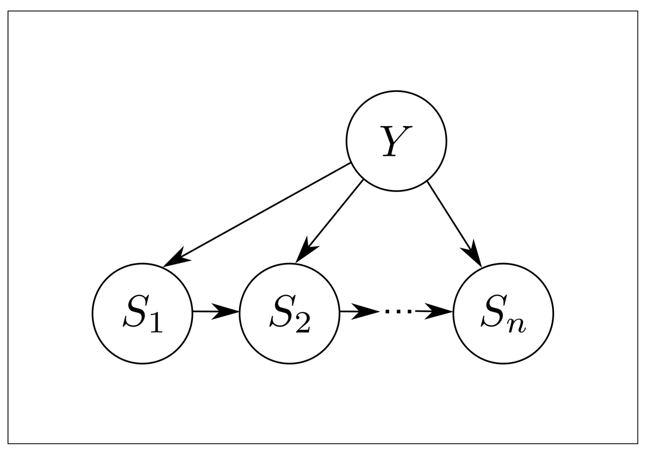
The proposed affordance-based object recognition system consists of the following subsystems:

- *information processing:* raw sensor information is fed into the *sensory processing* module (clustered GIST features) and is subsequently stored in the *sensorimotor feature* (SMF) alongside interaction information.
- *knowledge representation:* provides a learned *sensorimotor representation* in the form of a joint distribution of SMF and object class Y.
- *probabilistic reasoning:* uses a Bayesian network to infer the object class and provides a new interaction command obtained by an information gain strategy.

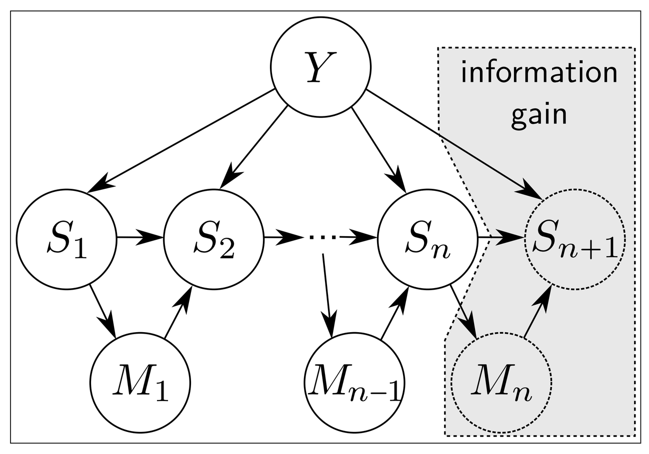
Bayesian Inference / Information Gain

We designed two types of Bayesian networks (BN) which process different kinds of information.

- The *Sensor Network (BN1)* processes only the sensor information with an extended naive Bayes approach which additionally allows for statistical dependencies between the preceding (s_{i-1}) and the current (s_i) sensor information. The information gain strategy can not be employed as no interaction information is available.
- The *Affordance-based Network (BN2)* processes the whole information stored in an SMF by assuming that the current sensor information s_i depends on the interaction m_i and the preceding sensor information (s_{i-1}). Additionally, *BN2* allows for statistical dependencies between the interaction m_i and the the preceding sensor information (s_{i-1}). The integration of interaction information allows to use an information gain strategy to choose an optimal next interaction m^*.



BN1
$P(y|s_{1:n}) \propto P(y)P(s_1|y)\prod_{i=2}^{n} P(s_i|s_{i-1}, y)$



BN2
$P(y|s_{1:n}, m_{1:n-1})$
$\propto P(y)P(s_1|y)\prod_{i=2}^{n} P(s_i|s_{i-1}, m_{i-1}, y)P(m_{i-1}|s_{i-1})$

The information gain IG of a possible next action m_n is defined as the difference between the current entropy and the conditional entropy:

$$IG(m_n) := H(Y|s_{1:n}, m_{1:n-1}) - H(Y|S_{n+1}, m_n, s_{1:n}, m_{1:n-1})$$

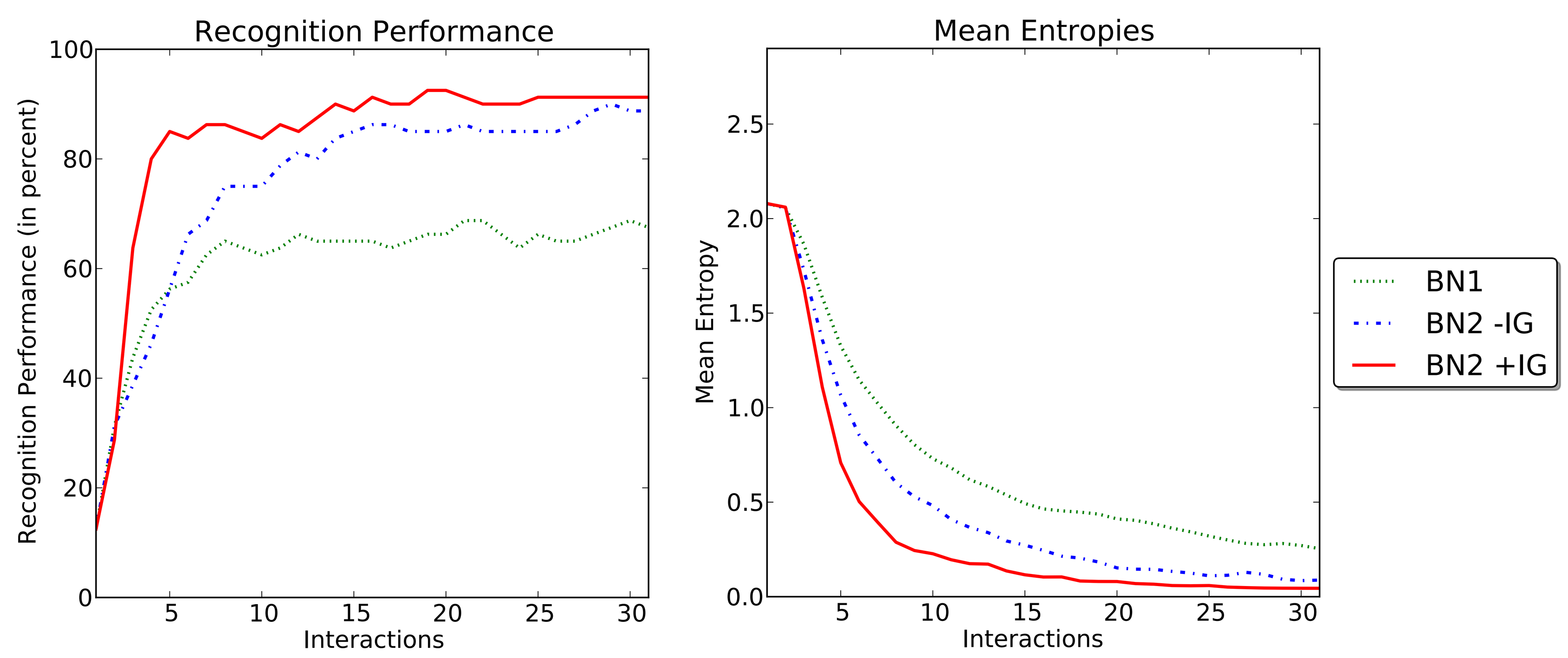
This is equivalent to the mutual information of Y and (S_{n+1}, m_n) for an arbitrary m_n. As the current entropy $H(Y|s_{1:n}, m_{1:n-1})$ is independent of the next action m_n the most promising action m^* can be calculated by minimizing the expected entropy with respect to S_{n+1}:

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

Evaluation and Results

Ten fold cross validation was conducted on a data set made by a robotic arm with a camera attached. The dataset has the following properties:

- 8 Object classes
- 10 objects in each class
- 435 SMFs per object
- 30 absolute positions
- 95 possible relative movements
- 30 interactions conducted



Conclusion

- The integration of affordance-based interaction results in better recognition performance.
- The information gain strategy leads to the acquisition of relevant information with fewer interactions.

Universität Bremen
Deutsches Zentrum für Luft- und Raumfahrt DLR
DFG Deutsche Forschungsgemeinschaft
UNI FREIBURG