FP6-004381-MACS

MACS

Multi-sensory Autonomous Cognitive Systems Interacting with Dynamic Environments for Perceiving and Using Affordances

Instrument: Specifically Targeted Research Project (STReP)

Thematic Priority: 2.3.2.4 Cognitive Systems

D5.2.1 Implementation of unsupervised and reinforcement learning algorithms

Due date of deliverable: August 31, 2005
Actual submission date: October 26, 2005

Start date of project: September 1, 2004 Duration: 36 months

Joanneum Research (JR_DIB)

Revision: Version 1

Project co-funded by the European Commission within the Sixth Framework Programme (2002–2006)		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

EU Project MACS

Deliverable D5.2.1

Learning of Affordance Perception

Lucas Paletta, Gerald Fritz, and Christin Seifert

Number: **MACS/5/2/1**
WP: *5.2*
Status: *draft, version 1*
Created at: *September 20, 2005*
Revised at: *October 26, 2005*

Learning of Affordance Perception

FhG/AIS	Fraunhofer Institut für Autonome Intelligente Systeme, Sankt Augustin, D
JR_DIB	Joanneum Research Graz, A
LiU-IDA	Linköping Universitet, Linköping, S
METU-KOVAN	Middle East Technical University, Ankara, T
OFAI	Österreichische Studiengesellschaft für Kybernetik,Vienna, A

This research was funded by the European Commission's 6th Framework Programme IST Project MACS under contract/grant number FP6-004381. The Commission's support is gratefully acknowledged.

Author addresses:

Lucas Paletta, Gerald Fritz, and Christin Seifert
Joanneum Research
Institute of Digital Image Processing
Computational Perception (CAPE)
Steyrergasse 9
A-8010 Graz, Austria

Fraunhofer Institut für
Autonome Intelligente Systeme
Schloss Birlinghoven
D-53754 Sankt Augustin
Germany

Tel.: +49 (0) 2241 14-2683
(Co-ordinator)

Contact:
Dr.-Ing. Erich Rome

Joanneum Research
Institute of Digital Image Processing
Computational Perception (CAPE)
Steyrergasse 9
A-8010 Graz
Austria

Tel.: +43 (0) 316 876-1769

Contact:
Dr. Lucas Paletta

Linköping Universitet
Dept. of Computer and Info. Science
Linköping 581 83
Sweden

Tel.: +46 13 24 26 28

Contact:
Prof. Dr. Patrick Doherty

Middle East Technical University
Dept. of Computer Engineering
Inonu Bulvari
TR-06531 Ankara
Turkey

Tel.: +90 312 210 5539

Contact:
Prof. Dr. Erol Sahin

Österreichische Studiengesellschaft
für Kybernetik (ÖSGK)
Freyung 6
A-1010 Vienna
Austria

Tel.: +43 1 5336112 0

Contact:
Prof. Dr. Georg Dorffner

Contents

1 Executive Summary

This deliverable report sets the background for the usage of reinforcement learning methods in the frame of affordance perception, and presents first experimental results about reinforcement learning from visual cues.

We first discriminate in Section 2 between (i) *(un)supervised learning* and (ii) *reinforcement learning* with respect to the problems given in the context of learning affordance cueing and recognition. The second step is then outlined in Section 3 towards a definition of perceptual states from visual cues that underlie any utility driven reasoning in observable and not observable Markovian decision processes. Finally, Section 4 refers to the important problem of perceptual aliasing, and how it refers to the notion of affordance object classes. This will open completely new insights into the understanding of 'object' definition in the context of functional object recognition.

This report is supposed to represent a 'living document', providing a first starting point which will become augmented from developments on learning affordance cues and affordance recognition during the course of the MACS project.

2 Learning of Affordance Cueing and Recognition

2.1 Learning vs. Heuristic Definition

A particular affordance must be recognized from an agent so that he becomes capable to decide upon interaction with the environment with respect to the corresponding affordance.

Affordance cues can either be heuristically defined or can be learned from repeated trials and presentations of a cue-to-(behaviour,outcome) association being exposed to the camera of the observing agent. Heuristic definition of affordance cues requires complete preconception of possible perceptual state trajectories within the task, determination of the perceptual state space, and inability to adjust for inconsistencies in the visual representation. Examples for heuristic definitions of cuntional object recognition are found in Deliverable D3.1.4.

In contrast, learning the affordance cues enables to select the triggering perceptual state according to the minimization of cost functions that are in direct context of the goal-determining task. The key task in machine learning based extraction of affordance cues is to estimate the class of vectors in visual feature space that enables reliable prediction of outcomes under the constraint of an interactive behaviour bridging the affordance cue to the behaviour outcome.

Figure 1 sketches the relation between affordance cueing, recognition and rewarding in an overview diagram. For the decision making agent, the purpose in finding affordance cues lies in the opportunity to be capable to decide early for a complete behavior on the basis of perceptual cues, *before any interaction with the physical world is necessary.* in this sense, the agent can act appropriately on the availability of affordances in the environment. Once the behavior is selected, it is processed until a final (terminal, goal, outcome) state is clearly reached. Feedback in the form of reward signals are either issued when the goal state is reached, or, according to an on-going distribution in reaction to particular state transitions.

An affordance cue in this sense is defined by *the perceptual cue that specifies a potential state transition (in a statistical sense) to a (behaviour, outcome) pair.* Affordance cues are

'per se' not determined to refer to any data structure with multimodal features. Instead, affordance cues might be patterns of distance relations, simple color or texture cues, visually perceivable object parts, etc., i.e., affordance cues are just any set or configuration of features that is distinguishable in their capability to predict the (behaviour,outcome) pair. Since we restrict the space of visual representations under investigation to the subspace of descriptors in vector space, we refer affordance cueing to the hierarchical vector hierarchy defined for MACS-specific representations (cf. Deliverable D3.1.1).

A detailed description about the learning of affordance recognition is outlined in Deliverable D5.3.1.

2.2 Supervised vs. Reinforcement Learning

Reinforcement learning [Sutton and Barto, 1998] is learning what to do – how to map situations to actions– so as to maximize a numerical reward signal. The learner is not told which actions to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward but also the next situation and, through that, all subsequent rewards. These two characteristics–*trial-and-error search* and *delayed reward* –are the two most important distinguishing features of reinforcement learning. Such an agent must be able to sense the state of the environment to some extent and must be able to take actions that affect the state. The agent also must have a goal or goals relating to the state of the environment. The formulation is intended to include just these three aspects–sensation, action, and goal–in their simplest possible forms without trivializing any of them.

Reinforcement learning is different from **supervised learning**, the kind of learning studied in most current research in machine learning, statistical pattern recognition, and artificial neural networks. Supervised learning is learning from examples provided by a knowledgable external supervisor. This is an important kind of learning, but alone it is not adequate for learning from interaction. In interactive problems it is often impractical to obtain examples of desired behavior that are both correct and representative of all the situations in which the agent has to act. In uncharted territory–where one would expect learning to be most beneficial–an agent must be able to learn from its own experience.

We applied **decision tree** based learning described in Deliverable D3.1.2 (see Section 5.2) for the purpose of estimating the function values in $\varphi : (\mathcal{P}, \mathcal{B}) \mapsto \mathcal{O}$, with percept vector $\mathcal{P}$, behaviour representationj $\mathcal{B}$, and outcomes $o \in \mathcal{O}$. Here, we assumed full awareness of a point in time where we determined the percept vector $\mathcal{P}$ at avery selected region within the captured image, and where it would make sense to investigate from this its direct relation to the single subsequent action a (i.e. in general, behaviour $\mathcal{B}$) and the outcome $\emptyset$. Actually, we are not allowed in general to assume existence of this knowledge, but we have to try out via multiple trials where to look at, i.e., where to sample the visual features in order to become capable from that investigation to find out any associative relationship between these predictive cues and the later event of an affordance application event (Deliverable D5.3.1). Any method that should be capable of finding the time or configuration where to look at, has to be capable of backtracking in time (or through state space) the information of the later affordance realisation back to the early stages of perception. With other words, it must be capable of solving a delayed reward problem, which is actually the case in reinforcement learning.

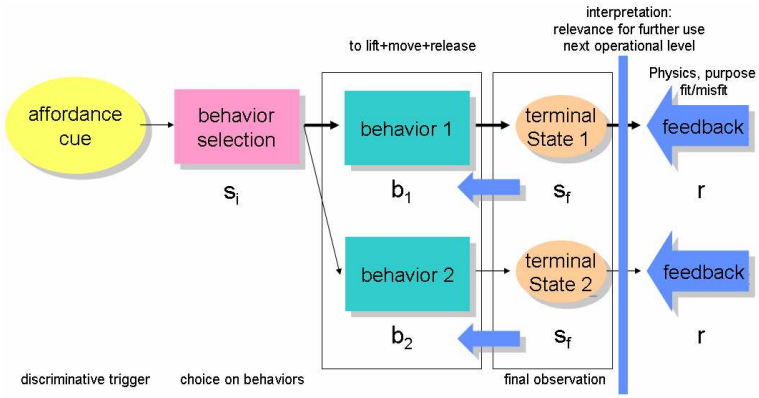


Figure 1: Affordance cueing, behaviour selection, and reinforcement signals for the learning of affordance cues.

3 Reinforcement Learning from Visual Cues

In Section 3.1 we outline in more detail the context of reinforcement learning, affordance behaviors, and perceptual states. In Section 3.2, we demonstrate successful learning from visual cues in a concrete example of learning spatial attention patterns in an object recognition task.

Summarising, reinforcement learning of recognition from visual cues requires reasonable structuring of the recognition task, in order to focus the decision process task on the very core of the action selection problem. The example shows that this can be done in ann efficient and feasible way.

3.1 Reinforcement Learning of Affordance Behaviors and Perception

Visual information provides per se complex cues about the environment. Furthermore, in typical visual perception tasks as is the case in object recognition, the task of object information generates multiple degrees of freedom (DOF) - six DOF in its position in space, in addition it suffers from variation in illumination and scale. Therefore, and originally, computer vision tasks impose highly complex problems on robustness and optimisation, and, methodologies to solve these tasks should not suffer from the curse of dimensionality.

Reinforcement learning is derived from dynamic programming (DP) which suffers in a way from the curse of dimensionality. In DP, the number of states grows exponentially with the number of state variables. On the other hand, reinforcement learning has already been successfully applied to problems with millions of states [Sutton and Barto, 1998]. Therefore, and inparticular with respect to computer vision tasks, careful design of the decision process is requested.

Figure 2 demonstrates the principle of structuring vision tasks in a natural way [Paletta et al., 2005a]. The complete process of state acquisition, action selection and reward computation is considered to take place in three major information processing stages as follows

- **Early vision.** The crucial operation in this stage is the reduction of the complete visual information to a restricted set of functional responses. It is helpful to apply

top-down processing taking advantage of contextual information for the selection of the local information, such as, in the informative features approach [Fritz et al., 2005b]. Here, the pixel information is summarised by local descriptor responses (e.g., using the Scale Invariant Feature Transform – SIFT [Lowe, 2004]). In the subsequent selection stage, only those descriptors are retained that provide high information value with respect to the task goal, i.e., discrimination among a set of object classes. This provides a highly converging number of relevant descriptors, and provides high accuracy in the repeatability of this selected information.

- **Feature Coding.** In this processing stage discretization can be performed on the focus of attention (FOA), i.e., the unit region of observation that was selected in the first stage. Here it is crucial to apply a transformation of the original information onto a low-dimensional representation that will determine the complexity of the state definition, and at the same time the feasability of the reinforcement learning task, thereafter. E.g., in [Paletta et al., 2005b], it is recommended to identify a small set of reference vectors that represent the complete distribution of samples that individually describe a particular instance of a FOA region.

- **Attention Control.** This is a characteristic reinforcement learning stage that can take advantage from the strucutring and selection of information during the previous processing stages. From top to bottom: First, a state representation is generated from the codebook vectors produced in stage 2, but also from any additional context knowledge, such as, information about geometrical relations between succeeding descriptor locations as in [Paletta et al., 2005b]. The next stage is to relate this state information to the task goal which can be defined by internally or externally generated reward signals. Here, we internally produce the signal since in pure object recognition, the success signal must be related to the internal model of the world and the task. Finally, the reinforcement learning agent - basically a decision maker form realised within a Markov Decision Process task (MDP or POMDP [Sutton and Barto, 1998]) - receives input in terms of state and reward information, applies the utility function estimator for any possible action, and from this generates a choice on action selection. Note that these actions might be directed either towards the usage of external action application, changing the configuration of the physical world, such as in wheel motor activities, or, the action might be applied within an internal choice on the selection of component selection [Whitehead and Ballard, 1991]. However, a strategy will be learned from multiple trials in order to improve the accuracy of the utility function estimator and the accummulation of reward during the task under investigation.

This concept can not only be applied in the learning of saccade driven attention patterns as outlined in Section 3.2 but will also be a framework for the learning of affordance cues and the recognition of affordances.

We propose here to use an affordance recognition component to decide about whether a goal state of the affordance completion behaviour (see Deliverable D3.1.2) has already been attained, or, otherwise to estimate the distance to it. This would determine a reward signal in analogy to how reward signmals are generated in robot navigation tasks [Sutton and Barto, 1998], namely, by issueing reward exactly at the moment when the goal state has been reached, and zero otherwise. It is guaranteed by the DP framework and the way

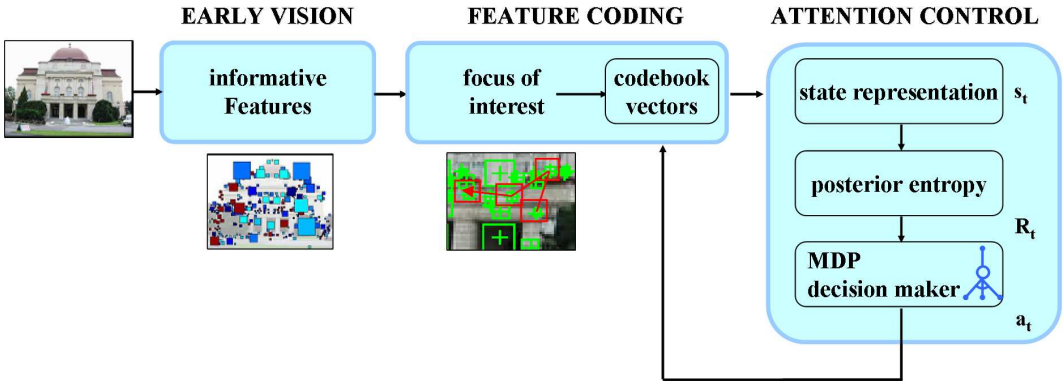


Figure 2: Concept of structuring the visual perception process for the application of reinforcement learning (here for the case of learning attention patterns for object recognition). In early vision, the system extracts informative local descriptors and focus of interest, where descriptors are encoded with respect to codebook vectors. Descriptor-action sequences define the state, posterior and entropy decrease to drive useful actions – closing the loop.

how utility function estimators are updated during the trial phase, that the reward signal is then backpropagated to sattes that are situated earlier in the typical sequence of states within the affordance completion behaviour.

In the following Section we describe in more detail how reinforcmeent learning is applied using visual cues for sequential attention tasks. The corresponding software can be used in MACS for the fast identification of regions or obejcts that are relevant for the cueing of affordances.

3.2 Reinforcement Learning of Visual Attention Patterns

Visual attention is crucial for autonomous systems in general [Paletta et al., 2005d], and in particular, within the MACS project for the early selection of regions of interest (ROI), but also - as described here - for the selective integration of information in a goal driven vision task as it is the case in affordance cueing.

We describe here the sequential attention concept for the task of object recognition, the methodology of strategy acquisition via reinforcement learning from trial and error, and illustrate it with first experimental results.

Attention is a highly important and emerging phenomenon in infant development [Ruff and Rothbart, 1996; Paletta et al., 2005c]. In human perception, sequential visual sampling of the environment is mandatory for object recognition purposes. Recent research in neuroscience [Schall and Thompson, 1999; Deco, 2004] and experimental psychology [Gorea and Sagi, 2003; Henderson, 2003; Deubel, 2004] has confirmed evidence that decision behavior plays a dominant role in human selective attention in object and scene recognition . E.g., there is psychophysical evidence that human observers represent visual scenes not by extensive reconstructions but merely by purposive encodings via saccadic attention patterns [Stark and Choi, 1996; Rybak et al., 1998] of few relevant scene features. This leads on the one hand to the assumption of transsaccadic object memories [Deubel, 2004], and supports theories about the effects of sparse in-

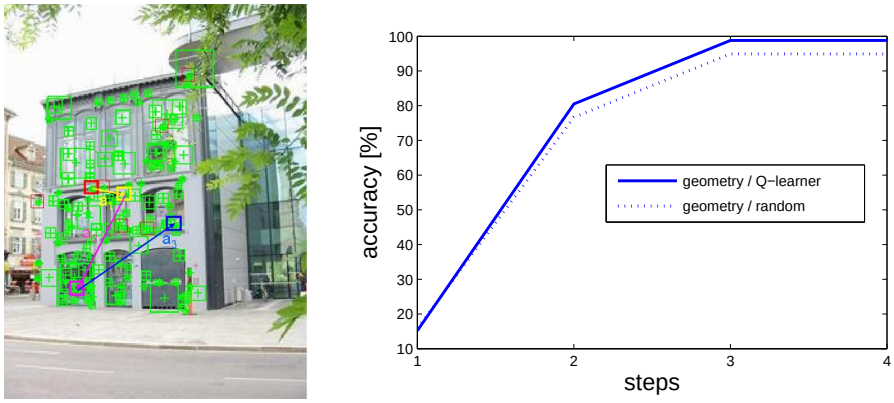
formation sampling due to change blindness when humans are caused to re-build visual interpretation under impact of attentional blinks [Rensink et al., 1997]. Current biologically motivated computational models on sequential attention identify shift invariant descriptions across saccade sequences [Li and Clark, 2004], and reflect the encoding of scenes and relevant objects from saccade sequences in the framework of neural network modeling [Rybak et al., 1998] and probabilistic decision processes [Bandera et al., 1996; Minut and Mahadevan, 2001].

In computer vision, recent research has been focusing on the integration of information received from single local descriptor responses into a more global analysis with respect to object recognition [Weber et al., 2000; Lowe, 2004]). State-of-the-art solutions, such as, (i) identifying the MAP hypothesis from probabilistic histograms [Fritz et al., 2004], (ii) integrating responses in a statistical dependency matrix [Weber et al., 2000], and (iii) collecting evidence for object and view hypotheses in parametric Hough space [Lowe, 2004], provide convincing performance under assumptions, such as, statistical independence of the local responses, excluding segmentation problems by assuming single object hypotheses in the image, or assuming regions with uniformly labelled operator responses. An integration strategy closing methodological gaps when above assumptions are violated should therefore (i) cope with statistical dependency between local features of an object, (ii) enable to segment multiple targets in the image and (iii) provide convincing evidence for the existence of object regions merely on the geometry than on the relative frequency of labelled local responses.

The original contribution of this work [Paletta et al., 2005b] is to provide a scalable framework for cascaded sequential attention in real-world environments. Firstly, it proposes to integrate local information only at locations that are relevant with respect to an information theoretic saliency measure. Secondly, it enables to apply efficient strategies to group informative local descriptors using a decision maker. The decision making agent used Q-learning to associate *shift of attention*-actions to cumulative reward with respect to a task goal, i.e., object recognition. Objects are represented in a framework of perception-action, providing a transsaccadic *working* memory that stores useful grouping strategies of a kind of *hypothesize and test* behavior.

In object recognition terms, this method enables to match not only between local feature responses, but also taking the geometrical relations between the specific features into account, thereby defining their more global visual configuration. The proposed method is outlined in a perception-action framework, providing a sensorimotor decision maker that selects appropriate saccadic actions to focus on target descriptor locations. The advantage of this framework is to become able to start interpretation from a single local descriptor and, by continuously and iteratively integrating local descriptor responses 'on the fly', being capable to evaluate the complete geometric configuration from a set of few features.

The saccadic decision procedure is embedded in a cascaded recognition process (Fig. 2) where visual evidence is probed exclusively at salient image locations. In a first processing stage, salient image locations are determined from an entropy based cost function on object discrimination. Local information in terms of code book vector responses determine the recognition state in the Markov Decision Process (MDP). In the training stage, the reinforcement learner performs trial and error search on useful actions towards salient locations within a neighborhood, receiving reward from entropy decreases. In the test stage, the decision maker demonstrates feature grouping by matching between the encountered



(a) (b)

Figure 3: Results of reinforcement learning of attention patterns: (a) A learned discriminative sequence of descriptor-action pairs that lead to a discriminative state with respect to the internal belief model on object hypotheses.

and the trained saccadic sensorimotor patterns.

3.2.1 Sensorimotor Sequential Attention

Sequential attention shifts the focus of attention between the most informative patterns in the order of associated saliency values, in this sense representing a step-wise generation of a scanpath [Stark and Choi, 1996]. There is two kind of information that characterizes objects for discrimination from scanpaths, (i) the visual information within the focus of attention, and (ii) the geometry between the sequentially accessed FOIs, i.e., the shift-of-attention action translating between FOIs. In this work we claim that the pattern in the FOI must not necessarily be represented in finest detail but an approximate characterization will suffice to give a weak object hypothesis. This renders the algorithm tolerant to noise and failures in the local interpretation, but on the other hand gives rise to analyse the spatial context, i.e., the geometry between the descriptors, in more detail.

Descriptor encodings The visual information in the FOI is associated to a prototypical reference vector to give a weak object hypothesis: At each local maximum, the extracted local pattern $\mathbf{g}_i$ is associated to a codebook vector Γ_j of nearest distance $d = \arg min_j ||\mathbf{g}_i - \Gamma_j||$ in feature space. The codebook vectors can be estimated from k-means clustering of a training sample set $G = \mathbf{g}_1, \cdots, \mathbf{g}_N$ of size N ($k = 20$ in the experiments). The focused local information pattern is therefore associated to the label of the k-th prototype vector, gaining discrimination merely from the geometric relations between focus encodings in order to discriminate attention patterns.

Action The shift-of-attention actions target in the proposed method towards one out of next n best-ranked maxima within the information theoretic saliency map. Saccadic actions originate from a randomly selected local maximum of saliency and target towards one of the remaining (n-1) best-ranked maxima via a saccadic action $a \in A$. The individual action and its corresponding angle $\alpha(x, y, a)$ is then categorized into one out of $|A| = 8$ principal directions ($\Delta a = 45°$).

Scanpath An individual state s_i is finally represented by a complete (or part of) a sequential attention pattern, i.e., the scanpath. The attention pattern of length n is

encoded by a sequence of descriptor encodings Γ_j and actions $a \in A$, i.e.,

$$s_i = (\Gamma_1, a_2, \cdots, \Gamma_{n-1}, a_n, \Gamma_n)\,. \tag{1}$$

Posteriors In order to characterize the discriminative value of a scanpath, we determine an estimate on the posterior on object hypotheses, given a particular descriptior-action sequence. The posterior is estimated from frequency histogramming: Within the object learning stage, random actions will lead to arbitrary descriptor-action sequences, i.e., attention patterns. For each attention pattern, we protocol the number of times it was experienced in the context of the corresponding object in the database. From this we are able to estimate a mapping from states s_i to posteriors, i.e., $s_i \mapsto P(o_k|s_i)$, by monitoring how frequent states are visited under observation of particular objects. From the posterior we compute the conditional entropy $H_i = H(O|s_i)$ and the *information gain* with respect to actions leading from state $s_{i,t}$ to $s_{j,t+1}$ by

$$\Delta H_{t+1} = H_t - H_{t+1}. \tag{2}$$

An efficient strategy aims then at selecting in each state $s_{i,t}$ the action a^* that would maximize the information gain $\Delta H_{t+1}(s_{i,t}, a_{k,t+1})$ received from attaining state $s_{j,t+1}$, i.e.,

$$a^* = \arg max_a \Delta H_{t+1}(s_{i,t}, a_{k,t+1}). \tag{3}$$

3.2.2 Q-Learning of Attentive Saccades

In each state of the sequential attention process (see Section 3.2.1), a decision making agent is asked to perform a strategy to select an action to arrive at a most reliable recognition decision. Learning to recognize objects means then to explore different descriptor-action sequences, to quantify consequences in terms of a utility measure, and to adjust the control strategy thereafter. In the following we motivate to define sequential attention as a decision process, and address to use reinforcement learning to extract the optimal policy from explorative search since we lack a precise model of the underlying statistics.

Markov decision processes (MDPs) have already been introduced for object recognition by [**?**] in the sense of optimal selection of visual procedures. Here, the MDP will provide the general framework to outline sequential attention for object recognition in a multistep decision task with respect to the discrimination dynamics. An MDP is defined by a tuple $(\mathcal{S}, \mathcal{A}, \delta, \mathcal{R})$ with state recognition set $\mathcal{S}$, action set $\mathcal{A}$, probabilistic transition function δ and reward function $\mathcal{R} : \mathcal{S} \times \mathcal{A} \mapsto \Pi(\mathcal{S})$ describes a probability distribution over subsequent states, given the attention shift action $a \in \mathcal{A}$ executable in state $s \in \mathcal{S}$. In each transition, the agent receives reward according to $\mathcal{R} : \mathcal{S} \times \mathcal{A} \mapsto R$, $\mathcal{R}_t \in R$. The agent must act to maximize the utility $Q(s, a)$, i.e., the expected discounted reward

$$Q(s,a) \equiv U(s,a) = E\left[\sum_{n=0}^{\infty} \gamma^n \mathcal{R}_{t+n}(s_{t+n}, a_{t+n}))\right], \tag{4}$$

where $\gamma \in [0, 1]$ is a constant controlling contributions of delayed reward.

We formalize a sequence of action selections $a_1, a_2, \cdots, a_n$ in sequential attention as an MDP and are searching for optimal solutions with respect to finding action selections so as to maximizing future reward with respect to the object recognition task. With each

shift-of-attention, the entropy reduction gives feedback about the reduction of uncertainty and therefore the quality of a related recognition decision. With each action, the reward in terms of information gain (Eq. 2) in the posterior distribution on object hypotheses, is received from attention shift a by

$$\mathcal{R}(s,a) := \Delta H. \tag{5}$$

Since the probabilistic transition function $\Pi(\cdot)$ cannot be known beforehand, the probabilistic model of the task is estimated via reinforcement learning, e.g., by Q-learning [Watkins and Dayan, 1992] which guarantees convergence to an optimal policy applying sufficient updates of the Q-function $Q(s,a)$, mapping recognition states s and actions a to utility values. The Q-function update rule is

$$Q(s,a) = Q(s,a) + \alpha \left[R + \gamma(max_{a'}Q(s',a') - Q(s,a))\right], \tag{6}$$

where α is the learning rate, γ controls the impact of a current shift of attention action on future policy returns.

The decision process in sequential attention is determined by the sequence of choices on shift actions at a specific focus of interest (FOI). The agent selects then the action $a \in \mathcal{A}$ with largest $Q(s,a)$, i.e.,

$$a_T = \arg max_{a'}Q(s_T,a'). \tag{7}$$

The method is evaluated in experiments on reference object databses, applying recognition using the reference COIL-20 (indoor imagery, [Murase and Nayar, 1995]) and the TSG-20 object (outdoor imagery [Fritz et al., 2005b; Fritz et al., 2005a], see Figure 3) database, proving the method being computationally feasible and providing rapid convergence in the discrimination of objects.

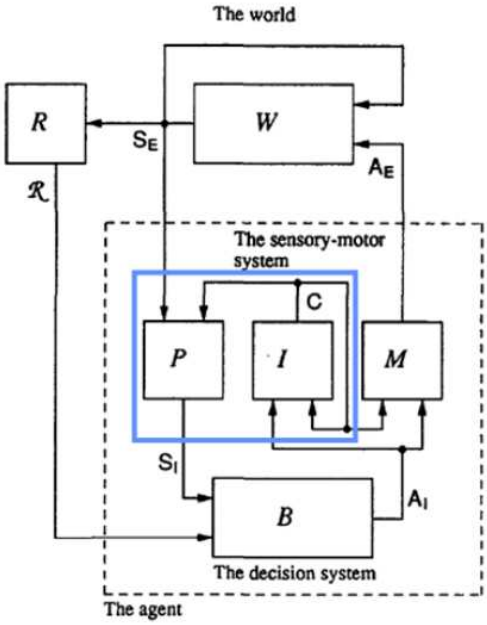
Summarising, we have shown via the presented method and experimental results that an efficient modeling of the visual information processing can make reinforcement learning feasible, and lay a basis for future modelling for the purpose of affordance cueing and recognition. Future work goes on cognitive system modeling using reinforcement learning tasks goes into the direction of applying contextual knowledge to prime the decision making agent on the object recognition task [Paletta et al., 2005e].

4 Perceptual Aliasing and Affordance Objects

4.1 Perceptual Aliasing in Reinforcement Learning

The use of perceptual cues for the definition of perceptual states in Markov decision processes (and reinforcement learning) leads to several design issues that need careful observation and treatment. Above all, visual features that are extracted locally in the image, in the most cases carry ambiguous information with respect to defining a unique state of perception (think of corner-like appearances that might be found at 100s of locations in the image). In reinforcement learning, the issue of finding distinctive perceptual states (in contrast to partially observable states) has been termed 'perceptual aliasing').

The potential of solving the problem of perceptual aliasing is linked in an interesting way to finding relevant features for the issue of affordance cueing: these issues are



Symbol	Function Name	Mapping	Probabilistic	Adaptive
W	Transition	$S_E \times A_E \rightarrow S_E$	Yes	No
R	Reward	$S_E \rightarrow \mathcal{R}$	Yes	No
P	Perceptual	$S_E \times C \rightarrow S_I$	No	No
I	Configuration	$A_I \times C \rightarrow C$	No	No
M	Motor	$A_I \times C \rightarrow A_E$	No	No
B	Behavioral	$(S_I \times \mathcal{R})^* \rightarrow A_I$	Yes	Yes

Figure 4: A formal model for an agent with an embedded learning subsystem and an active sensory-motor subsystem. The table summarizes the functions implemented by each of the model's modules (from Whitehead and Ballard (1991)).

identical in requesting to find appropriate state representations for disambiguating similar observations, and thereby, for appropriately predicting delayed consequences impacted by immediate actions.

[Whitehead and Ballard, 1991] separated a decision making agent into two subsystems, i.e., a decision and a perception subsystem (Figure 4), in order to illustrate the differences between external (SE) and internal (SI) state representations. The straightforward integration of visual cues for state representation and reinforcement learning can lead to undesirable interactions that prevent a decision subsystem from learning an optimal control strategy. These interactions arise because the mapping between world states and the agent's internal representation is many-to-many. That is, a state s in the world, depending upon the configuration of the sensory-motor subsystem, may map to several internal states; conversely, a single internal state, s', may represent multiple world states. This overlapping between the world and the agent's internal representation is called perceptual aliasing [Whitehead and Ballard, 1991].

Perceptual aliasing can have a devastating impact on the decision subsystem's ability to learn an adequate control policy because it causes the system to confound world states that it must necessarily distinguish in order to solve the task. Considering the effect of perceptual aliasing in a problem-solving tasks, one can find that two decision problems must be distinguished: the decision problem faced by the agent and the decision problem faced by the embedded decision subsystem. The problem faced by the agent is the original problem-solving task defined by the world. The problem faced by the decision subsystem corresponds to the original problem as transformed by the sensory-motor interface. [Whitehead and Ballard, 1991] call these the actual (or external) problem and the

perceived (or internal) problem, respectively.

A standard methodology to solve the perceptual aliasing problem is to solve partial observable Markov decision processes (POMDP) by searching the space of finite policies [Meuleau et al., 1999]. [Chrisman, 1992] proposed the predictive distinctions approach to compensate for perceptual aliasing caused from incomplete perception of the world. A model must be learned in addition to the control policy. The system must discover– on its own – the important distinctions in the world. [McCallum and Andrew, 1995] added with the approach of 'instance-based' learning of hidden states that are matched to history sequences. Instead of recording instances in a continuous geometrical space, one has to record instances in action-percept-reward sequence space. More recently, and in application to computer vision problems, [Jodogne and Piater, 2005] learned sets of image descriptors to disambiguate, and [Kuipers and Beeson, 2002] presented place recognition by integrating history to develop distinctive states.

4.2 Affordance Objects

In order to detect the visual features that are actually relevant for affordance cueing we intend to implement the approach by [Chrisman, 1992] and will compare it to the instance-based learning approach by [McCallum and Andrew, 1995]. We expect that the model learned in the predictive distinctions approach should exactly contain those feature groupings that enable to successfully predict the recognition of affordances, under the assumption of a particular behaviour in terms of a state-action sequence had taken place.

Understanding affordance cues in the context of determining a perceptual state, the issue of disambiguating states with perceptual aliasing can be seen as finding exactly those states, i.e., affordance cues, that are distinctive, i.e., relevant, for the prediction of consequences of behaviours. In particular, learning of these distinctive states, i.e., affordance cues, must be performed in an incremental manner, by starting with most general descriptions of perceptual states, and refining this state representation iteratively until no improvement will be gained anymore from state representation updates. The task in MACS is to guide this process of disambiguating in a way that enables the natural definition of more complex features, groupings, and, eventually, object data structures, until all hidden states are covered.

We will first implement reinforcement learning within a MACS-like scenario, and then start a sequence of experiments where we will drive the system to investigate to find increasingly complex groupings in order to improve performance measres, such as, accuracy of detection and identification, robustness against noise, and to reduce the comlpexity in processing and representation. It is planned to perform this work during the second year of the MACS project.

References

[Bandera et al., 1996] Bandera, C., Vico, F., Bravo, J., Harmon, M., and III, L. B. (1996). Residual Q-learning applied to visual attention. In *International Conference on Machine Learning*, pages 20–27.

[Chrisman, 1992] Chrisman, L. (1992). Reinforcement learning with perceptual aliasing: The perceptual distinctions approach. In *Proc. National Conference on Artificial Intelligence, AAAI-92*.

[Deco, 2004] Deco, G. (2004). The computational neuroscience of visual cognition: Attention, memory and reward. In *Proc. International Workshop on Attention and Performance in Computational Vision*, pages 49–58.

[Deubel, 2004] Deubel, H. (2004). Localization of targets across saccades: Role of landmark objects. *Visual Cognition*, (11):173–202.

[Fritz et al., 2004] Fritz, G., Paletta, L., and Bischof, H. (2004). Object recognition using local information content. In *Proc. International Conference on Pattern Recognition, ICPR 2004*, volume II, pages 15–18. Cambridge, UK.

[Fritz et al., 2005a] Fritz, G., Seifert, C., and Paletta, L. (2005a). Urban object recognition from informative local features. In *Proc. IEEE International Conference on Robotics and Automation, ICRA 2005*, pages 132–138, Barcelona, Spain.

[Fritz et al., 2005b] Fritz, G., Seifert, C., Paletta, L., and Bischof, H. (2005b). Learning informative sift descriptors for attentive object detection. In *Proc. Joint Hungarian-Austrian Conference on Image Processing and Pattern Recognition, HACIPPR 2005*, pages 95–102, Veszprm, Hungary. Received Best Paper Award of AAPR 2005.

[Gorea and Sagi, 2003] Gorea, A. and Sagi, D. (2003). Selective attention as the substrate of optimal decision behaviour in environments with multiple stimuli. In *Proc. European Conference on Visual Perception*.

[Henderson, 2003] Henderson, J. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Sciences*, 7:498 –504.

[Jodogne and Piater, 2005] Jodogne, S. and Piater, J. (2005). Interactive learning of mappings from visual percepts to actions. In *Proc. International Conference on Machine Learning, ICML 2005*.

[Kuipers and Beeson, 2002] Kuipers, B. and Beeson, P. (2002). Bootstrap learning for place recognition. In *Proc. National Conference on Artificial Intelligence, AAAI-02*.

[Li and Clark, 2004] Li, M. and Clark, J. (2004). Learning of position and attention-shift invariant recognition across attention shifts. In *Proc. International Workshop on Attention and Performance in Computational Vision*, pages 41–48.

[Lowe, 2004] Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

[McCallum and Andrew, 1995] McCallum, R. and Andrew, R. (1995). Instance-based utile distinctions for reinforcement learning. In *Proc. International Conference on Machine Learning, ICML 1995*, Lake Tahoe, CA.

[Meuleau et al., 1999] Meuleau, N., Kim, K.-E., Kaelbling, L., and Cassandra, A. (1999). Solving pomdps by searching the space of finite policies. In *Proc. International Conference on Uncertainty in Artificial Intelligence, UAI 1999*.

[Minut and Mahadevan, 2001] Minut, S. and Mahadevan, S. (2001). A reinforcement learning model of selective visual attention. In *Proc. International Conference on Autonomous Agents*, pages 457–464.

[Murase and Nayar, 1995] Murase, H. and Nayar, S. K. (1995). Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision*, 14(1):5–24.

[Paletta et al., 2005a] Paletta, L., Fritz, G., and Seifert, C. (2005a). Perception-action based object detection from local descriptor combination and reinforcement learning. In *Proc. 19th Scandinavian Conference on Image Analysis, SCIA 2005*, LNCS 3540, pages 639–648, Joensuu, Finland. Springer-Verlag, Berlin Heidelberg.

[Paletta et al., 2005b] Paletta, L., Fritz, G., and Seifert, C. (2005b). Q-learning of sequential attention for visual object recognition from informative local descriptors. In *Proc. 22nd International Conference on Machine Learning, ICML 2005*, pages 649–656, Bonn, Germany.

[Paletta et al., 2005c] Paletta, L., Fritz, G., and Seifert, C. (2005c). Reinforcement learning of informative attention patterns for object recognition. In *Proc. 4th International Conference on Development and Learning, ICDL 2005*, pages 188–193, Osaka, Japan.

[Paletta et al., 2005d] Paletta, L., Rome, R., and Buxton, H. (2005d). *Neurobiology of Attention*, chapter Attention Architectures for Machine Vision and Mobile Robots, pages 642–648. Academic Press, New York, NY.

[Paletta et al., 2005e] Paletta, L., Seifert, C., and Fritz, G. (2005e). Contextual working memory for trans-saccadic object recognition using reinforcement learning and informative local descriptors. In *Proc. European Conference on Visual Perception, ECVP 2005*, page 69, A Coruna, Spain. Abstract.

[Rensink et al., 1997] Rensink, R., O'Regan, J., and Clark, J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8:368–373.

[Ruff and Rothbart, 1996] Ruff, H. and Rothbart, M. (1996). *Attention in early development*. Oxford University Press, New York, NY.

[Rybak et al., 1998] Rybak, I., V., I. G., Golovan, A., Podladchikova, L., and Shevtsova, N. (1998). A model of attention-guided visual perception and recognition. *Vision Research*, 38:2387–2400.

[Schall and Thompson, 1999] Schall, J. and Thompson, K. (1999). Neural selection and control of visually guided eye movements. *Annual Review of Neuroscience 22:*, (22):241–259.

[Stark and Choi, 1996] Stark, L. W. and Choi, Y. S. (1996). Experimental metaphysics: The scanpath as an epistemological mechanism. In Zangemeister, W. H., Stiehl, H. S., and Freska, C., editors, *Visual attention and cognition*, pages 3–69. Elsevier Science, Amsterdam, Netherlands.

[Sutton and Barto, 1998] Sutton, R. S. and Barto, A. (1998). *Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning)*. MIT Press, Cambridge.

[Watkins and Dayan, 1992] Watkins, C. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3,4):279–292.

[Weber et al., 2000] Weber, M., Welling, M., and Perona, P. (2000). Unsupervised learning of models for recognition. In *Proc. European Conference on Computer Vision*, pages 18–32.

[Whitehead and Ballard, 1991] Whitehead, S. and Ballard, D. (1991). Learning to perceive and act by trial and error. *Machine Learning*, 7(1):45–83.